Banco de México

Documentos de Investigación

Banco de México

Working Papers

N° 2011-08

Time-series Modelling, Stationarity and Bayesian Nonparametric Methods

Juan Carlos Martínez-Ovando
Banco de México

Stephen G. Walker
University of Kent

September 2011

Documento de Investigación
2011-08

Working Paper
2011-08

Time-series Modelling, Stationarity and Bayesian Nonparametric Methods*

Juan Carlos Martínez-Ovando†
Banco de México

Stephen G. Walker‡
University of Kent

Abstract: In this paper we introduce two general non-parametric first-order stationary time-series models for which marginal (invariant) and transition distributions are expressed as infinite-dimensional mixtures. That feature makes them the first Bayesian stationary fully non-parametric models developed so far. We draw on the discussion of using stationary models in practice, as a motivation, and advocate the view that flexible (non-parametric) stationary models might be a source for reliable inferences and predictions. It will be noticed that our models adequately fit in the Bayesian inference framework due to a suitable representation theorem. A stationary scale-mixture model is developed as a particular case along with a computational strategy for posterior inference and predictions. The usefulness of that model is illustrated with the analysis of Euro/USD exchange rate log-returns.
Keywords: Stationarity, Markov processes, Dynamic mixture models, Random probability measures, Conditional random probability measures, Latent processes.
JEL Classification: C11, C14, C15, C22, C51.

Resumen: En este artículo introducimos dos modelos generales no paramétricos estacionarios de primer orden para series de tiempo, para los cuales las distribuciones marginales (invariantes) y de transiciones son expresadas como mezclas de dimensión infinita. Esta característica los hacen los primeros modelos Bayesianos estacionarios completamente no paramétricos desarrollados a la fecha. Contribuimos a la discusión sobre el uso de modelos estacionarios en la práctica, como una motivación, y apoyamos la visión que, modelos estacionarios flexibles (no paramétricos) pueden ser una fuente confiable para inferencias y predicciones. Se observa que nuestros modelos se acoplan apropiadamente al esquema Bayesiano de inferencia debido a un adecuado teorema de representación. Un modelo estacionario de mezcla de escalas es desarrollado como caso particular junto con una estrategia computacional para realizar inferencia posterior y predicciones. La utilidad de este modelo es ilustrado con el análisis de los log-retornos del tipo de cambio euro-dólar E.U.A.
Palabras Clave: Estacionariedad, Procesos de Markov, Modelos de mezclas dinámicos, Medidas aleatorias de probabilidad, Medidas aleatorias de probabilidad condicional, Procesos latentes.

*The authors thank two anonymous referees for their constructive comments and suggestions. The first author gratefully acknowledges financial and other support by CONACYT (Mexico), University of Kent (UK), Regione Piemonte, University of Turin and Collegio Carlo Alberto (Italy).

†Dirección General de Investigación Económica. Email: juan.martinez@banxico.org.mx. (Corresponding author).

‡School of Mathematics, Statistics and Actuarial Science. Email: S.G.Walker@kent.ac.uk.

1 Introduction

Much of statistical theory as we know it relies on some sort of symmetry among the observations (see, e.g. Zabell, 2005). In the context of time-series analysis the notion of symmetry has been expressed in terms of stationarity, from which large sample properties and other theoretical results are derived. However, a drawback that most of the stationary models have exhibited, which make practitioners doubt about their usefulness, has been exposed by the restrictive structural nature of either their transition mechanisms or marginal distributions. It is our thesis that the use of flexible stationary models can produce sensible inferences and predictions. In what remains of this section we shall review some general flaws found in stationary models when dealing with real data and comment on some approaches that people have found to overcome the lack of flexibility of existing stationary models. Then, we shall argue that stationary models when connected with Bayesian nonparametric methods can be as flexible as many of their non-stationary counterparts, or even more so, while preserving all the desired properties that a statistical model may have. We draw now our attention to the connection between stationarity and Bayesian nonparametric methods.

1.1 Stationarity and Bayesian nonparametrics

The world in which we live is exposed to permanent and sometimes unanticipated changes. So are the data that are collected from it. Due to unanticipated changes observed in data, particularly in the time-series context, practitioners have developed a hesitance about the use of stationary models. In the Econometric literature, an influential theory to forecasting, which accounts to fix some failures of existing parametric econometric models, can basically be characterized by using mixtures of models whose structural forms do not really differ in more than some specific values of the parameters involved. See Clements and Hendry (2008) for a recent overview. Let us notice that those models are often specified as non-stationary. However, they combine different mechanisms describing the transition dynamics of the data and (possibly) anticipating future values and sudden changes. See also Krolzig (1997) and Frühwirth-Schnatter (2006), where a Markovian dynamic structure is imposed to describing the data transitions from one model component to another. Note that Markov switching

models were proposed by Hamilton (1989) to modelling non-stationary patterns of economic time-series data. The relevance and good performance in terms of prediction/forecasting of different types of mixture models in empirical econometrics has been demonstrated in Clements and Hendry (1999), even though these models suffer the lack of theoretical support that stationary models can provide. Similarly, non-stationary models have been used in Financial Econometrics and Financial Statistics (see, e.g. Giacominia et al., 2008), with the aim of capturing the dynamic mechanism of the data generating process in a flexible way, if such a mechanism exists. See also Haas et al. (2004) and Barry and Winkler (1976) where some implications of using non-stationary models in portfolio analysis are discussed.

Related to the problem of modelling structural changes in Statistics, attention has also been put on developing statistical procedures aiming at identifying (and testing) whether observed changes in the data represent actual structural changes to the model. See, e.g. Chu et al. (1996) and Zeileis et al. (2005), for a discussion. But, as it is well acknowledged, statistical procedures aiming at identifying structural breaks suffer the flaw of just identifying structural changes retrospectively and not much can be said about future changes. Therefore, it would be ideal to conceive a statistical model that integrates different structural possibilities into a single specification, as it will capture a large variety of scenarios of the data. In a way, that has been the aim of general mixtures and Markov-mixtures models existing in the literature.

Preserving the idea of integrating different structural forms into a single model, it is reasonable to believe that infinite-dimensional mixture transition dynamics would be a sensible and adequate alternative to modelling the transition dynamics of time-series data. Such type of models would potentially capture persistent characteristics of the data, as well as transitory events, but the model will also leave open the possibility of considering alternative scenarios for future inferences and, more importantly, predictions. An infinite-dimensional model would ideally learn and put heavy probabilities to recurrent events and lower to rare observed events. An additional desirable feature that infinite-dimensional transition mechanics may have, and that non-stationary models may not be able to do, is that of reproducing the marginal behaviour of the data for any time, which has also been a flaw of most finite-dimensional mixture models developed so far. Then, Bayesian nonparametric methods may

come into play. Refer to Lijoi and Prünster (2010) for a recent overview. Some infinite-dimensional mixture models for time-series data have been recently developed, but so far none of them can reproduce the marginal behaviour of the data in a flexible way (see some of the papers included in Barber et al., 2008, 2011). What this paper adds to the literature, is the introduction of fully nonparametric stationary models admitting flexible marginal and transition distributions, which are represented as infinite-dimensional mixtures, making these models the unique existing alternatives achieving that goal within the Bayesian nonparametric framework. The model construction we adopt follows the latent processes approach of Pitt et al. (2002), but we exploit a more flexible stochastic structure around which the dependence of the stationary model is built.

1.2 About stationarity and representation theorems

In addition, within the Bayesian statistics framework, the notion of conditional stationarity and conditional ergodicity of a model becomes relevant as it gives rise to an analogous dynamic version of de Finetti's representation theorem (de Finetti, 1937). Such a result is actually set upon the more general notion of stochastic symmetry under invariant measures, due to Maitra (1977) (see also, Aldous, 1985). Maitra's representation states that for a countable sequence of stationary random variables, say $\{Y_t\}_{t=1}^{\infty}$, its associated probability measure, $\mathbb{P}$, admits the following representation for any finite sub-segment of the sequence of length n and any measurable sets $(A_t)_{t=1}^{n}$,

$$\mathbb{P}(Y_1 \in A_1, \ldots, Y_n \in A_n) = \int_{\Theta} \prod_{t=1}^{n} F_\theta(y_t \in A_t | y_{t-1}, \ldots, y_1) \Pi(\mathrm{d}\theta), \tag{1.1}$$

where θ is a random element taking values in the space Θ (which can be finite- or infinite-dimensional), Π is a probability measure with support in Θ, and the $F_\theta(y_t \in A_t | y_{t-1}, \ldots, y_1)$'s represent a collection of conditionally stationary and conditionally ergodic distributions indexed by θ, with $F_\theta(y_1 \in A_1)$ being the marginal distribution of the model at time $t = 1$. It is worth to notice that the representation (1.1) encompasses the de Finetti's representation theorem for exchangeable random variables, in which case conditional stationarity is replaced with conditional independence given θ. de Finetti's representation theorem has motivated the Bayesian subjective approach to inference and prediction under the assumption

of exchangeability.

A particular case to (1.1) is obtained when the F_θ's are Markov, in which case the whole model is characterized by marginal, $F_\theta(\cdot)$, and first-order transitions, $F_\theta(\cdot|\cdot)$. Just as in the exchangeable case, the probability measure Π would represents one's prior beliefs about the model. Other dependent models may also accommodate Maitra's representation. Hence, from a subjective point of view, Maitra's representation theorem can be seen as motivation for the development of subjective Bayesian inference for time-series analysis under stationarity. It is also noticeable that there is no representation result like (1.1) available for non-stationary models.

Returning our attention to (1.1), it is worth to notice that when choosing F_θ and $\Pi(\mathrm{d}\theta)$ one would automatically place a prior distribution on a given space of stationary models. For instance, when the F_θ's are parametric and θ is finite-dimensional, the model will be reduced to traditional parametric models. In that case, one would be assigning null probability to many large sets of alternative models (i.e. parametric models whose functional forms lie outside the class of models where the F_θ's belong), hence the posterior probability for those sets given the data will be null as well. Thus, it seems sensible to look for robust specifications of F_θ and $\Pi(\mathrm{d}\theta)$ which encompass a large class of models, in order to avoid surprises after observing the data. Again, Bayesian nonparametric specifications, where Θ is typically infinite-dimensional, seems to be an appealing alternative to such specifications.

In this paper, we introduce two constructive procedures which aim at defining fully non-parametric stationary Markov models admitting Maitra's representation. We anticipate that one of them, the benchmark model, cannot be implemented in practice with the computational machinery available. However, the second extended model overcomes that flaw. It is worth to point out that the extended models have alternative state-space representations. These models are appealing alternatives to traditional state-space models, as the observable part of these models can be recovered analytically due that observation and state components are regular conditional distributions. A stationary scale-mixture model is introduced as a particular case of the latter and a computational algorithm for posterior inference is developed. It is also worth to point out that the fully non-parametric stationary models introduced here can be adapted to a variety of circumstances, just by specifying some of

their components according to different needs (e.g., having different supports).

1.3 Layout of the paper

The layout of the paper is as follows. Section 2 presents a general review of first-order stationary modelling. There we introduce a benchmark model, which generalizes the nonparametric kernel type estimator of marginal and transition distributions of a Markov process; we also comment on a practical limitation. General ideas related to the latent processes model construction approach are also summarized. In Section 3 we introduce an extended framework to specifying first-order stationary models admitting nonparametric marginal and transition densities, which overcomes the limitation attained to the benchmark model. In Section 4, a particular stationary-scale mixture model is developed around double nonparametric mixtures of beta-Stacy and stick-breaking processes. In Section 5 we specify priors for the stationary scale-mixture model and develop upon analytic and computational strategies for posterior inference. The applicability of the stationary scale-mixture model is illustrated in Section 6 with the analysis of a short segment of the Euro/USD exchange rate log-returns. We conclude with some final remarks and comments in Section 7. The paper is complemented with appendices where a general overview of the latent processes model construction based upon mixtures is presented and details regarding the implementation of the dynamic scale-mixture model are given.

2 First-order stationary models

2.1 Previous approaches

Some approaches for constructing first-order stationary time-series models pay attention to specifying an explicit analytic form for the marginal distribution compatible with the transition mechanism of the model. Among them, we find the approaches based on convolutions (Jacobs and Lewis, 1978) and the ones based on latent processes (Pitt et al., 2002).

The convolution-based approach of Jacobs and Lewis (1978) makes use of the thinning operator to defining stationary discrete-valued time-series models with given marginals. Gaver

and Lewis (1980) explored that idea and defined a stationary model with gamma marginals. Joe (1996) exploited the property of closeness under convolution of the exponential dispersion family of distribution functions and defined stationary models with marginals in the family of convolution closed exponential family. Related extensions have been further studied by Jørgensen and Song (1998), Jørgensen and Song (2006), and Zhu and Joe (2006), among others. More recent generalizations considering the binomial thinning operator have also been studied by Bouzar and Jayakumar (2008) and Bouzar (2010).

The latent-process approach was developed by Pitt et al. (2002). Their model construction focused on models admitting marginals in the conjugate exponential and exponential dispersion families of distribution functions, while preserving linearity of the conditional expectations of the transition dynamics. Further extensions admitting more structured marginals and transition dynamics were later developed by Pitt and Walker (2005, 2006), in which stationary versions of ARCH and GARCH models (see, e.g. Engle, 2001) with given marginals were also developed. The scope of the latent-process approach was further extended by Mena and Walker (2007) by considering latent-random probability measures, which might give rise to more structured (flexible) transition mechanisms while preserving a given marginal distribution. That idea was applied in Contreras-Cristán et al. (2009) where alternative stationary discrete-valued time-series models are defined.

However, despite all the flexibility that both approaches to model construction may have, so far the marginal (invariant) distributions they give rise still belong to a given parametric family of distribution functions. In this paper we will introduce two more flexible first-order stationary models for which both marginal and transition distribution are non-parametric. Let us start with the specification of what we term a benchmark model.

2.2 A benchmark model

First-order stationary models are specified by marginal and second-order joint distributions. Generically speaking, we can consider an observable bivariate stochastic component, (Y_{t-1}, Y_t), whose entries take values in the common space $\mathcal{Y}$. Additionally, we can assume that any pair (Y_{t-1}, Y_t) follows a joint symmetric finite-dimensional distribution, which is denoted by $K(y_{t-1}, y_t|\theta)$, depending upon a parameter θ, such that the following relation is

satisfied

$$p(y_t|\theta) = \int p(y_t|y_{t-1};\theta)p(y_{t-1}|\theta)\mathrm{d}y_{t-1}, \tag{2.1}$$

where $p(y_t|y_{t-1};\theta) = K(y_{t-1}, y_t|\theta)/p(y_{t-1}|\theta)$ with $p(\cdot|\theta)$ being the common marginal distribution induced by K. The parameter θ would take values in the space Θ.

A natural Bayesian nonparametric extension to the joint density for (Y_{t-1}, Y_t) would be given in terms of an infinite-dimensional mixture model, in the sense of Lo (1984), which may take the form

$$p(y_{t-1}, y_t) = \sum_{j=1}^{\infty} W_j K(y_{t-1}, y_t|\theta_j). \tag{2.2}$$

The bivariate distribution would then be characterized by the weights $\{W_j\}_{j=1}^{\infty}$ and the parameters $\{\theta_j\}_{j=1}^{\infty}$, whose elements share a finite-dimensional parameter space Θ. As a consequence, the common marginal distribution induced by (2.2) would be given by

$$\begin{aligned} p(y_t) &= \sum_{j=1}^{\infty} W_j \int K(\mathrm{d}y_{t-1}, y_t|\theta_j) \\ &= \sum_{j=1}^{\infty} W_j K(y_t|\theta_j). \end{aligned} \tag{2.3}$$

Hence, a nonparametric version of the first-order transition density would take the form

$$\begin{aligned} p(y_t|y_{t-1}) &= \frac{p(y_{t-1}, y_t)}{p(y_{t-1})} \\ &= \frac{\sum_{j=1}^{\infty} W_j K(y_{t-1}, y_t|\theta_j)}{\sum_{l=1}^{\infty} W_l K(y_{t-1}|\theta_l)}, \end{aligned} \tag{2.4}$$

for any t.

Before proceeding, let us point out that the transition mechanism of the benchmark model (2.4) resembles –and extends in terms of flexibility– the kernel type estimator for the transition dynamics of a Markov process with general state space that was originally proposed by Roussas (1969), which has served as a benchmark model for nonparametric estimation of Markov processes in the frequentist framework. Over the years, kernel type estimators for Markov processes have received significant attention. See, for example, Athreya and Atuncar (1998), Hili (2001), Campos and Dorea (2005) and Lacour (2008).

But notice that despite the simplicity of the expression (2.4) and the fact of being a flexible generalization to traditional kernel type estimators of the transition distribution, its

practical tractability is limited due to the issue involving an infinite sum in its denominator. Therefore, the computational implementation of the benchmark model is not possible, for the moment. However, in the next section, we derive a flexible generalization which overcomes the intractability of the denominator involved in similar representations to (2.4). The stationary models we derive there admit infinite-dimensional mixture representations for the marginal and transition, simultaneously. That is achieved by looking at the transition distribution $p(y_t|y_{t-1})$ as a mixture, in the spirit of Pitt et al. (2002), rather than as a quotient, as above. We refer to such a model construction approach as the Gibbs sampler model construction.

2.3 The Gibbs sampler model construction

The Gibbs sampler model construction framework was originally developed by Pitt et al. (2002), and it was addressed in a broader framework by Mena and Walker (2005). The construction relies on the incorporation of latent probability measures as an instrument to induce a desired dependence structure and marginal behaviour of the model. Let us continue by establishing some notation. Let Y be an observable random variable taking values in the sample space $\mathcal{Y}$, and let f be a density (or probability mass function) for Y, which is assumed to be random. The set-up is completed with the incorporation of a probability measure Π for f, which is defined on the infinite-dimensional function space $\mathcal{F}$ where f belongs. Accordingly, the model construction is derived from the following structure

$$y|f \sim f, \quad \text{and} \quad f \sim \Pi. \tag{2.5}$$

As it is often assumed, the probability measure Π is defined in such a way that the relation $\mathbb{E}_\Pi(f) = g_0$ holds, for any given density function g_0. Hence, it is straightforward to see that the marginal density for Y would be given by g_0, i.e.

$$p(y_t) = \int f(y_t)\Pi(\mathrm{d}f) = \mathbb{E}_\Pi[f(y_t)] = g_0(y_t). \tag{2.6}$$

The key idea of the Gibbs sampler model construction consists in expressing the transition density from time $(t-1)$ to t as a mixture of the form

$$p(y_t|y_{t-1}) \quad = \quad \int f(y_t)\Pi(\mathrm{d}f|y_{t-1}) = \mathbb{E}_\Pi(f(y_t)|y_{t-1}), \tag{2.7}$$

for which explicit analytic expressions exist for almost any probability measure Π known in the literature (assuming that Y_{t-1} and Y_t are conditionally independent given f; see, e.g. Lijoi and Prünster, 2010). Also, under suitable specifications of Π, it will be possible to find flexible representations for the transition dynamics, (2.7), as studied by Mena and Walker (2005).

A potential drawback of the existing Gibbs sampler model construction is that the form of the marginal density, (2.6), is typically parametric. Thus, the possibility of defining a fully nonparametric alternative to the benchmark model described above seems not viable. However, a general approach which circumvent that drawback is developed in Section 3. The idea consists in extending the basic Gibbs sampler model construction in two directions. On the one hand, we shall explore suitable double mixture representations for Π, which allow us to express the marginal density as an infinite-dimensional mixture. On the other hand, we shall define a suitable joint observable-latent representation of the transition dynamics of the process, which will allow us to circumvent the infinite-sum involved in the denominator of the transition distribution of the benchmark model. Hence, the transition mechanism of the model would be written as an infinite-dimensional mixture.

Before proceeding, let us notice that in the heart of the Gibbs sampler model construction framework there lies the specification of a joint probability measure induced by f and Π, say $\mathbb{P}$, which is given by

$$\mathbb{P}(\mathrm{d}y, \mathrm{d}f) = f(y)\Pi(\mathrm{d}f)\mathrm{d}y. \tag{2.8}$$

The support of $\mathbb{P}$ is given by the product space $\mathcal{Y}\times\mathcal{F}$. The notion of $\mathbb{P}$ will be fundamental for the model construction we shall describe in Section 3, as the desired marginal and transition mechanisms will be derived from alternative structured versions of (2.8).

But before entering in the details of the extended Gibbs sampler model construction we will stablish a remarkable connection with the benchmark model. It happens, as we shall see below, that the benchmark model can be derived as a particular case of Gibbs sampler stationary models.

2.3.1 Derivation of the benchmark model via the Gibbs sampler

The benchmark model emerges as a particular case of the Gibbs sampler model construction when $K(y_{t-1}, y_t|\theta_j) = K(y_{t-1}|\theta_j)K(y_t|\theta_j)$ and the follow joint probability measure on the product space $\mathcal{Y} \times \mathcal{J}$ is defined

$$\mathbb{P}(\mathrm{d}y_t, J = d) = W_d K(\mathrm{d}y_t|\theta_d), \tag{2.9}$$

where $\mathcal{Y}$ is the sample space of Y and $\mathcal{J} = \{1, 2, \ldots\}$. The role that the latent variable J plays in this construction is that of a latent imputation or latent index variable associated with a mixture model. Therefore, from (2.9) it follows that the conditional probability measure of $J = d$, given y_{t-1}, would be

$$\mathbb{P}(J = d|y_{t-1}) \propto W_d K(y_{t-1}|\theta_d), \tag{2.10}$$

with constant of proportionality given by $\sum_{l=1}^{\infty} W_l K(y_{t-1}|\theta_l)$. Hence, imputing (2.10) into the analogous version of (2.7) gives rise to the alternative Bayesian nonparametric kernel type transition mechanism

$$\begin{aligned} p(y_t|y_{t-1}) &= \sum_{d=1}^{\infty} \mathbb{P}(y_t|d)\, \mathbb{P}(d|y_{t-1}) \\ &= \frac{\sum_{d=1}^{\infty} W_d K(y_t|\theta_d) K(y_{t-1}|\theta_d)}{\sum_{l=1}^{\infty} W_l K(y_{t-1}|\theta_l)}. \end{aligned} \tag{2.11}$$

Hence, following this derivation, it is straightforward to see that the marginal stationary distribution associated with (2.11) would coincide with

$$p(y_t) = \sum_{l=1}^{\infty} W_l K(y_t|\theta_l), \tag{2.12}$$

for any $t = 1, 2, \ldots$, as desired. Let us proceed with the extended model construction.

3 Extended stationary dynamic mixture models

The extended Gibbs sampler model construction we shall develop overcomes the practical limitation that the benchmark model has, while representing marginal and transition mechanisms of the model via infinite-dimensional mixtures. The construction relies on an extended

representation of the joint probability measure $\mathbb{P}$, as in (2.8), but now incorporating a further latent probability measure, which will play a fundamental role in the description of the transition mechanism of observable and latent variables. A general derivation based upon mixtures is described in Appendix A.

3.1 Invariant distribution

Similarly to the benchmark model introduced above, we shall look for stationary models admitting a marginal (invariant) density as an infinite-dimensional mixture. The explicit form of such a distribution will be given here, after establishing some additional notation. Let Y be an observable random variable taking values in the measurable (Polish) space $(\mathcal{Y}, \mathscr{B}_{\mathcal{Y}})$. The stationary models we are interested in admit a marginal (invariant) density (or probability mass function, if required) of the form

$$p(y) = \sum_{j=1}^{\infty} W_j \int_{\mathcal{L}} K(y|\lambda) G_0(\mathrm{d}\lambda; \theta_j), \tag{3.1}$$

where $\{W_j\}_{j=1}^{\infty}$ is a sequence of probability weights, such that $0 \leq W_j \leq 1$, for any j, and $\sum_{j=1}^{\infty} W_j = 1$; λ is a mixing variable taking values in the measurable space $(\mathcal{L}, \mathscr{B}_{\mathcal{L}})$; $K(y|\lambda)$ is a parametric kernel defined on $\mathcal{Y} \times \mathcal{L}$; $G_0(\cdot|\theta_j)$ is a given distribution function defined on the measurable space $(\mathcal{L}, \mathscr{B}_{\mathcal{L}})$, which is indexed by the parameter θ_j, for each j; and $\{\theta_j\}_{j=1}^{\infty}$ is a sequence of parameters taking values in a given common (finite-dimensional) space Θ. It is also assumed that the probability distributions $G_0(\cdot|\theta_j)$'s have exactly the same functional form across j's, differing only in terms of the parameters θ_j's. It will be assumed that G_0 is a diffuse probability measure on $(\mathcal{L}, \mathscr{B}_{\mathcal{L}})$ with density function g_0.

As it is typically done with mixture models, the scope to (3.1) can be extended by means of introducing a latent imputation variable, J, which indicates from which component of mixture the observable random variable Y comes. In that case, the augmented version to (3.1), after including $J = d$ as latent variable, turns into

$$p(y, d) = W_d \int_{\mathcal{L}} K(y|\lambda) G_0(\mathrm{d}\lambda; \theta_d), \tag{3.2}$$

for any $d \in \mathcal{J} = \{1, 2, \ldots\}$.

3.2 Stochastic set-up

From the augmented set-up derived in (3.2), we can further assume that there exists a random distribution function, F, taking values in $\mathcal{F}$, which is defined on the measurable space $(\mathcal{L}, \mathscr{B}_{\mathcal{L}})$. It would also be assumed that for each j there is a probability measure indexed by θ_j, say $\Pi(\cdot;\theta_j)$, defined on the space $(\mathcal{F}, \mathscr{B}_{\mathcal{F}})$ such that $G_0(\cdot|\theta_j) = \mathbb{E}_{\Pi}\{F(\cdot)|\theta_j\}$. As before, G_0 is assumed to have the same functional form across j's, differing only in terms of the θ_j's. The probability measures $\Pi(\cdot;\theta_j)$ will have the same structural form, but this time differing only in terms of their corresponding baseline probability distributions $G_0(\cdot|\theta_j)$'s. The reason for explicitly including F in the stochastic set-up will become evident later in the paper. For the moment, let us anticipate that it will play a fundamental role in the specification of the transition dynamics of the model, as its transition mechanism will be defined around the F itself.

Hence, by means of expanding the scope of (3.2), after including F and λ as additional components, it would be possible to define an extended stochastic structure, $\mathbb{P}$, defined on the augmented product space $\mathcal{Y} \times \mathcal{L} \times \mathcal{F} \times \mathcal{J}$, given by

$$\mathbb{P}(\mathrm{d}y, \mathrm{d}\lambda, \mathrm{d}F, J = d) = K(\mathrm{d}y|\lambda)F(\mathrm{d}\lambda)\Pi(\mathrm{d}F|G_0(;\theta_d))W_d. \tag{3.3}$$

This stochastic structure may have different interpretations or derivations. We believe that among the possible alternatives, the next analogous hierarchical representation of (2.5) would be more representative in our context:

$$\begin{aligned} Y|\lambda \sim K(y|\lambda), \qquad & \lambda|F \sim F(\lambda), \\ F \quad \sim \quad & \widetilde{\Pi}(\cdot) = \sum_{j=1}^{\infty} W_j \Pi(\cdot|G_0(;\theta_j)), \end{aligned}$$

where it is explicitly seen that the chosen probability measure $\widetilde{\Pi}$ is represented as a double nonparametric mixture. See Teh et al. (2006) and Rodríguez et al. (2008) for related structured Bayesian nonparametric set-ups.

The above augmented probability measure will serve as the stochastic set-up for our model construction. The next section develops the transition dynamics of the model induced by (3.3). The starting framework to defining the transition dynamics is the augmented set-up in which the transition dynamics will be given in terms of the triplet of observable and

latent variables, (y, λ, d), and the dependence structure is to be defined around F through its associated probability measure Π.

3.3 Transition dynamics

In order to define the transition dynamics, we consider the state at time $(t-1)$ for the triplet of observable and latent variables $(y_{t-1}, \lambda_{t-1}, d_{t-1})$ which, given F, will have the joint probability measure

$$\mathbb{P}(\mathrm{d}y_{t-1}, \mathrm{d}\lambda_{t-1}, J_{t-1} = d_{t-1}|F) = W_{d_{t-1}} K(\mathrm{d}y_{t-1}|\lambda_{t-1}) F(\mathrm{d}\lambda_{t-1}). \tag{3.4}$$

Accordingly, the transition mechanism between $(t-1)$ and t would be characterized by the probability transition rule

$$\begin{aligned} \mathbb{P}(\mathrm{d}y_t, \mathrm{d}\lambda_t, d_t|y_{t-1}, \lambda_{t-1}, d_{t-1}) &= \int_{\mathcal{F}} \mathbb{P}(\mathrm{d}y_t, \mathrm{d}\lambda_t, d_t|F)\, \mathbb{P}(\mathrm{d}F|y_{t-1}, \lambda_{t-1}, d_{t-1}) \\ &= W_{d_t} K(\mathrm{d}y_t|\lambda_t) G(\mathrm{d}\lambda_t|\lambda_{t-1}, d_{t-1}), \end{aligned} \tag{3.5}$$

with

$$\begin{aligned} \mathbb{P}(\mathrm{d}F|y_{t-1}, \lambda_{t-1}, d_{t-1}) &= \Pi(\mathrm{d}F|\lambda_{t-1}, d_{t-1}), \\ G(\mathrm{d}\lambda_t|\lambda_{t-1}, d_{t-1}) &= \mathbb{E}_{\Pi}\{F(\mathrm{d}\lambda_t)|\lambda_{t-1}, d_{t-1}\}, \end{aligned} \tag{3.6}$$

where $\Pi(\mathrm{d}F|\lambda_{t-1}, d_{t-1})$ stands for the conditional probability measure of F given (λ_{t-1}, d_{t-1}), whereas $G(\mathrm{d}\lambda_t|\lambda_{t-1}, d_{t-1})$ stands for the one-step predictive rule under Π for λ_t conditionally on (λ_{t-1}, d_{t-1}).

3.3.1 Explicit expressions of conditional and one-step expected rules for F

Explicit expressions for the conditional probability measure for F given (λ_{t-1}, d_{t-1}) can be obtained according to the values of the latent index d_{t-1} and the values the latent index variable J_t may take. That is to say

$$\Pi(\mathrm{d}F|\lambda_{t-1}, d_{t-1}) \propto \begin{cases} \Pi(\mathrm{d}F|G_0(;\theta_{d_{t-1}}), \lambda_{t-1}) & , \quad j = d_{t-1}, \\ \Pi(\mathrm{d}F|G_0(;\theta_j)) & , \quad j \neq d_{t-1}, \end{cases} \tag{3.7}$$

where the constant of proportionality is defined in terms of the sequence $\{W_j\}_{j=1}^{\infty}$ and the remaining components involved in the model. Notice that the dependence of F upon λ_{t-1} will become explicit only in the case $d_t = d_{t-1}$. Otherwise, $\Pi(\mathrm{d}F|\lambda_{t-1}, d_{t-1})$ will get reduced to the marginal probability measure Π given the index value d_{t-1} only.

Accordingly, the one-step transition rule for the λ_t's will be given by

$$G(\mathrm{d}\lambda_t|\lambda_{t-1}, d_{t-1}) \propto \begin{cases} G_1(\mathrm{d}\lambda_t|\lambda_{t-1}, d_{t-1}; \theta_{d_{t-1}}) & , \quad j = d_{t-1}, \\ G_0(\mathrm{d}\lambda_t; \theta_j) & , \quad j \neq d_{t-1}, \end{cases} \tag{3.8}$$

where $G_1(\mathrm{d}\lambda_t|\lambda_{t-1}, d_{t-1}; \theta_{d_{t-1}}) = \mathbb{E}_{\Pi}\{F(\mathrm{d}\lambda_t)|\lambda_{t-1}, d_{t-1}; \theta_{d_{t-1}}\}$ corresponds to the one-step transition rule for λ_t under Π, conditionally on the pair (λ_{t-1}, d_{t-1}). It is worth to notice that the transition rule can be explicitly computed for most of the probability measures Π known in the literature, even in the case $d_t = d_{t-1}$ (see, e.g. Lijoi and Prünster, 2010).

3.3.2 Reduced form of the transition mechanism

Summarizing, it can be said that the dynamics of the sequence $\{(y_t, \lambda_t, d_t)\}_{t=1}^{\infty}$ are characterized by the transition rule:

$$\mathbb{P}(\mathrm{d}y_t, \mathrm{d}\lambda_t, d_t|y_{t-1}, \lambda_{t-1}, d_{t-1}) \propto \begin{cases} W_{d_{t-1}} K(\mathrm{d}y_t|\lambda_t) G_1(\mathrm{d}\lambda_t|\lambda_{t-1}, d_{t-1}; \theta_{d_{t-1}}) & , \quad d_t = d_{t-1}, \\ W_{d_t} K(\mathrm{d}y_t|\lambda_t) G_0(\mathrm{d}\lambda_t; \theta_{d_t}) & , \quad d_t \neq d_{t-1}, \end{cases} \tag{3.9}$$

with all the components involved given as before.

A distinctive feature of the transition dynamics of this model is that it is theoretically possible to recover the transition dynamics on the observables, $\{y_t\}$, requiring only the computation of

$$\begin{aligned} \mathbb{P}(\mathrm{d}y_t|y_{t-1}) &= \int \mathbb{P}(\mathrm{d}y_t|F)\, \mathbb{P}(\mathrm{d}F|y_{t-1}) \\ &= \int\int K(\mathrm{d}y_t|\lambda_t) F(\mathrm{d}\lambda_t)\, \mathbb{P}(\mathrm{d}F|y_{t-1}), \end{aligned} \tag{3.10}$$

where

$$\mathbb{P}(\mathrm{d}F|y_{t-1}) \propto \begin{cases} W_{d_{t-1}} \int \Pi(\mathrm{d}F|G_0(;\theta_{d_{t-1}}), \lambda_{t-1}) G_0(\mathrm{d}\lambda_{t-1}|y_{t-1}; \theta_{d_{t-1}}) & , \quad j = d_{t-1}, \\ W_j \Pi(\mathrm{d}F|G_0(;\theta_j)) & , \quad j \neq d_{t-1}, \end{cases}$$

with $G_0(\lambda_{t-1}|y_{t-1};\theta_{d_{t-1}}) \propto K(y_{t-1}|\lambda_{t-1})G_0(\lambda_{t-1}|\theta_{d_{t-1}})$.

It is verifiable that the pair (3.1) and (3.10) satisfy the balance equation of a Markov process. Hence, both marginal invariant and first order transition distribution define a first-order strictly stationary sequence $\{Y_t\}_{t=1}^{\infty}$.

Although it is possible to find an analytic expression for the transition rule (3.10), for inferential purposes, it will be sufficient to open the transition rule for observable and latent variables simultaneously. In this way, some statistical procedures developed to make inference on state-space models can be implemented here.

4 A stationary scale-mixture model

Here we develop a dynamic scale-mixture model as a particular case of the extended construction introduced in Section 3. The elements of the model would be given by the parametric kernel, which is taken to be a Gaussian, i.e. $K(y|\lambda) = \mathrm{N}(y|0,\lambda)$, where λ is a scale parameter. Hence, the state-space for λ is $\mathbb{R}_+$. As the baseline probability measure of Π we consider the Weibull distribution with shape parameter, $\theta > 0$, i.e. $G_0(\lambda|\theta) = \mathrm{We}(\lambda|\theta)$, with $\theta > 0$, and density function $g_0(\lambda|\theta) \propto \theta\lambda^{\theta-1}e^{-\lambda^{\theta}}$. The choice of G_0 as a Weibull distribution follows from the choice of Π, the probability measure for F, which is taken to be a beta-Stacy process (see, Walker and Muliere, 1997). The beta-Stacy process is chosen due the flexible structure of its one-step conditional expected distribution, and it will involve explicit expressions of the hazard function, h_0, associated with G_0. Choosing a parametric family for G_0 with explicit hazard functions is relevant for the model specification. Details regarding the beta-Stacy process are summarized below.

4.1 The beta-Stacy process

The beta-Stacy (BS) process was introduced by Walker and Muliere (1997) as a particular neutral-to-the-right process (Doksum, 1974). The attractive feature of the BS process, which is relevant for our model construction, is attributable to the flexibility that it is retained after computing one-step conditional expected distributions, i.e. G_1 according to the notation used. Let us summarize the definition and relevant characteristics of the beta-Stacy process.

A random probability distribution F, with support on the positive real line, is said to be a beta-Stacy process if it can be written as a suitable transformation of an increasing Lévy process, of the form

$$F(\lambda) \stackrel{d}{=} 1 - \exp\{-Z(\lambda)\}, \tag{4.1}$$

where $\{Z(\lambda) : \lambda \geq 0\}$ is a process with nondecreasing independent increments, such that: i) $Z(\lambda)$ is non-decreasing almost surely (a. s.), ii) $Z(\lambda)$ is right-continuous a. s., iii) $Z(0) = 0$ a.s., and iv) $\lim_{\lambda \to \infty} Z(\lambda) = +\infty$ a. s.. The beta-Stacy process arises when $Z(\lambda)$ is taken to be a log-beta process characterized by the non-homogeneous Lévy measure

$$\nu(\mathrm{d}u, \mathrm{d}s) = \frac{1}{1 - e^{-s}} e^{-\ s\beta(u)} \mathrm{d}s \ \alpha(\mathrm{d}u), \tag{4.2}$$

for $u > 0$ and $s \geq 0$, with α being a continuous finite measure on $(\mathbb{R}_+, \mathscr{B}_{\mathbb{R}_+})$ and β being a non-negative function on the positive real line, such that $\lim_{\lambda \to +\infty} \int_0^\lambda \frac{\alpha(\mathrm{d}u)}{\beta(u)} = +\infty$.

The particular parameterization that we consider for the BS process, which accommodates G_0 (the Weibull distribution) as the unconditional expected probability measure, is

$$\alpha(\mathrm{d}u) = \beta(u)\theta u^{\theta - 1} \mathrm{d}u, \tag{4.3}$$

with $\theta > 0$ and β being a positive function parameter with support on the positive real line.

Several attributes of the beta-Stacy process make it suitable for our model construction. Its structural conjugacy guarantees that if $F \sim \mathrm{BS}(\alpha, \beta)$ and $\lambda | F \sim F$, then the conditional probability measure for F given λ' (an exact value of λ) would also be a beta-Stacy process characterized by an updated log-beta process, Z^*, with a random jump of size S_1 at λ', such that

$$1 - \exp\{-S_{\lambda'}\} | \lambda' \sim \mathrm{Be}(1, \beta(\lambda')), \tag{4.4}$$

and a continuous component with updated Lévy measure ν^* given by

$$\nu^*(\mathrm{d}u, \mathrm{d}s) = \frac{1}{1 - e^{-s}} e^{-s[\beta(u) + \mathbb{1}(\lambda' \geq u)]} \mathrm{d}s \ \alpha(\mathrm{d}u). \tag{4.5}$$

See Corollary 2 in Walker and Muliere (1997) for further details.

Additionally, due to the fact that the process is being characterized by a non-homogeneous Lévy measure and the adopted parametrization, the beta-Stacy process has the additional

attractiveness of giving rise to flexible one-step conditional expected distributions which, conditionally on an exact observation λ', can be written as

$$\begin{aligned} G_1(\lambda|\lambda') &= \mathbb{E}_{\text{BS}}[F(\lambda)|\lambda'] = 1 - \mathbb{E}[e^{-Z^*(\lambda)}] \\ &= 1 - \exp\left\{ - \int_0^{\lambda} \left[\mu(u)\, \mathbb{1}_{(0,\lambda']}(u) + \mathbb{1}_{(\lambda',\infty)}(u)\right] h_0(u) \mathrm{d}u \right\} \\ &\quad \times \left[\mathbb{1}_{(0,\lambda')}(\lambda) + \mu(\lambda')\, \mathbb{1}_{[\lambda',\infty)}(\lambda)\right], \end{aligned} \tag{4.6}$$

with the latter expression derived after considering the re-parameterization $\mu(u) = \frac{\beta(u)}{\beta(u)+1}$. At the first glance, the functional form for G_1 seems be too intricate. But it actually is tractable, and represents a mixed-type distribution for λ which is characterized by: i) a functional form tempered (i.e. uniformly re-scaled) by μ over the region $(0, \lambda')$, ii) a point mass at λ', with probability mass depending upon μ, and iii) a right-tail behavior matching with G_0 over the region (λ', ∞). In order to make (4.6) computationally and analytically tractable it is convenient to consider a suitable parameterization for μ, which is described next.

4.2 A suitable parameterization for μ

Without loosing flexibility, we can further assume that the function parameter μ belongs to the class of piecewise and right-continuous $(0, 1)$-valued functions, with

$$\mu(u) = \sum_{k=0}^{\infty} \mu_k\, \mathbb{1}_{E_k}(u), \tag{4.7}$$

for $u \geq 0$, where $\{\mu_k\}_{k=0}^{\infty}$ is a $(0, 1)$-valued sequence and $\{E_k\}_{k=0}^{\infty}$ is a monotonic partition of the positive real line. By monotonic partition we refer to a partition induced by an increasing monotonic sequence $0 = \xi_0 < \xi_1 < \ldots < \xi_k < ...$, with $E_k = (\xi_k, \xi_{k+1}]$, for $k \geq 0$. Similar parameterizations of function parameters have previously been considered in Arjas and Gasbarra (1994) and Nieto-Barajas and Walker (2002), in different contexts.

Additionally, we introduce a further latent indicator variable, z, around which the density/probability mass function g_1, attained to G_1, can be written as the following mixture

$$g_1(\lambda|\lambda') = \sum_{z=-1}^{1} g_{1,z}(\lambda|\lambda') p(z|\lambda'), \tag{4.8}$$

where

$$z|\lambda' \sim \begin{cases} -1 & \text{with } p(z=-1|\lambda') = P\{(0,\lambda')|\lambda'\}) \\ 0 & \text{with } p(z=0|\lambda') = P(\{\lambda'\}) \\ 1 & \text{with } p(z=1|\lambda') = P\{(\lambda',\infty)|\lambda'\}. \end{cases} \tag{4.9}$$

Explicit analytic expressions for the relevant components involved in (4.8) are given in Appendix B.

4.3 The likelihood function and data augmentation scheme

Given a collection of observations $\boldsymbol{y} = \{y_t\}_{t=1}^T$, the likelihood function for the dynamic stationary scale-mixture model would involve components according to whether the two consecutive latent indexes d_t and d_{t-1} are equal or different. Specific components of the likelihood associated with the cases $d_t = d_{t-1}$ are difficult to handle analytically. Thus, a data augmentation scheme is required. In particular, we incorporate the indicator variables involved in the specification of g_1 as additional latent variables. Hence, the augmented likelihood after incorporating the z_t's would be written as

$$\begin{aligned} lik(Q,\mu,\boldsymbol{\lambda},\boldsymbol{z},\boldsymbol{d};\boldsymbol{y}) \;\propto\; & \left\{ \prod_{\{t:d_t=d_{t-1}\}} W_{d_t}\, \mathrm{N}(y_t|0,\lambda_t) g_{1,z_t}(\lambda|\lambda_{t-1};\theta_{d_t},\mu) p(z_t|\lambda_{t-1};\theta_{d_t},\mu) \right\} \\ & \cdot \left\{ \prod_{\{t:d_t\neq d_{t-1}\}} W_{d_t}\, \mathrm{N}(y_t|0,\lambda_t)\, \mathrm{We}(\lambda_t|\theta_{d_t}) \right\}, \end{aligned} \tag{4.10}$$

where $Q = \sum_{j=1}^{\infty} W_j \delta_{\theta_j}$, which is characterized by the joint sequence $\{(W_j,\theta_j)\}_{j=1}^{\infty}$, and μ is the piece-wise $(0,1)$-valued function characterized by the join sequence $\{(\xi_k,\mu_k)\}_{k=1}^{\infty}$. Additionally, $\boldsymbol{\lambda} = (\lambda_t)_{t=1}^{\infty}$ corresponds to the scale-latent process, and $\boldsymbol{d}$ corresponds to the latent sequence of index variables. Recall that explicit expressions for $p(z_t|\lambda_{t-1};\theta_{d_t},\mu)$ and $g_{1,z_t}(\lambda_t|\lambda_{t-1};\theta_{d_t},\mu)$ are given in Appendix B. What is worth to notice in the augmented likelihood (4.10), with regards to the latent component $\boldsymbol{z} = (z_t)$, is that in the computational implementation of the model the dimension of $\boldsymbol{z}$ changes across iterations according to the cases in which two consecutive latent index variables coincide. Let us proceed with a description of the prior-to-posterior analysis.

5 Posterior analysis

5.1 Stochastic truncation via latent variables

A necessary step in order to take the stationary scale-mixture model into practice consists in dealing with the infinite-sum still involved in the extended likelihood (4.10) for each d_t. Similarly to the approach used by Walker (2007) for infinite-dimensional mixture models, we introduce a latent-truncation variable to restrict the range of each d_t to a finite set. But, differently to Walker (2007), we do not use the slice sampler. Instead, we directly truncate the range of each d_t via a geometric latent variable, U_t.

Abbreviating notation, the component of the likelihood (4.10) involving each d_t may be written as

$$f(\lambda_t, d_t) = W_{d_t} f(\lambda_t|\theta_{d_t}), \tag{5.1}$$

for $d_t = 1, 2, \ldots$. The specific form of $f(\lambda_t|\theta_{d_t})$ in our context will depend on all the possible combinations of the index sets (d_{t-1}, d_{t+1}), in conjunction of d_t. [But notice that in a traditional mixture model setting, $f(\lambda_t|\theta_{d_t})$ would be expressed just by the kernel component of the mixture.] The stochastic truncation takes place in (4.10), at each time t, through (5.1) by means of incorporating the latent-truncation variable, U_t, such that its conditional distribution, given d_t, would be a shifted geometric random variable, i.e.

$$U_t|d_t \sim \phi(1-\phi)^{u-d_t}\, \mathbb{1}(u \geq d_t), \tag{5.2}$$

with $0 < \phi < 1$. It can be seen that the marginal distribution for U_t will be geometric with parameter ϕ. Consequently, the augmented version of (5.1), after introducing u_t, will be

$$f(\lambda_t, d_t, u_t) = W_{d_t} f(\lambda_t|\theta_{d_t})\phi(1-\phi)^{u_t-d_t}\, \mathbb{1}(u_t \geq d_t). \tag{5.3}$$

It is also verifiable that the original model (5.1) can be recovered from (5.3) after integrating out U_t.

Therefore, the stochastic truncation is implemented within the posterior Gibbs sampler where sampling from the conditional distributions for each d_t given u_t, and vice-versa, is required. Details about the further augmented likelihood, after incorporating the sequence

of additional latent truncating variables $\boldsymbol{u} = \{u_t\}_{t=1}^{T}$, and the implementation of the posterior Gibbs sampler are given in the appendices.

5.2 Prior specification

Here we describe two prototype Bayesian nonparametric priors for the two model parameters involved, namely μ and Q. It is further assumed that μ and Q are independent.

5.2.1 Prior for μ

Recall that μ, or the sequence $\{(\xi_k, \mu_k)\}_{k=1}^{\infty}$, can be seen as a realized path of a jump-process, which we assumed is driven by a regular homogeneous Poisson-marked process, whose probability law admits the decomposition, $\mathbb{P}\{(\xi_k, \mu_k)\} = \Pi\{(\xi_k)\} \cdot \Pi\{(\mu_k)|(\xi_k)\}$, with $\Pi\{(\xi_k)\}$ being the probability law of the homogeneous-Poisson process part with intensity $\nu_\xi > 0$; and $\Pi\{(\mu_k)|(\xi_k)\}$ being the probability law of the marked part, which is specified in terms of the following system of conditional probabilities

$$p(\mu_0|\xi_0) = \mathrm{Be}(\mu_0|a_\mu, b_\mu), \tag{5.4}$$

$$p(\mu_k|\xi_k, \ldots, \xi_0, \mu_{k-1}, \ldots, \mu_0) = \mathrm{Be}\left(\mu_k|a_\mu + k, c_\mu\right), \tag{5.5}$$

for $k = 1, 2, \ldots$. Accordingly, the sequence of random marks will be increasing in mean. In this prior specification the parameter c_μ regulates how fast we could expect the function μ will go to 1. The smaller the c_μ the slower the function μ will grow in mean. A similar prior specification on the space of piecewise-constant random functions was used by Arjas and Gasbarra (1994) to estimate random hazard functions in survival analysis.

5.2.2 Prior for Q

Recall that Q is characterized by the sequence $\{(W_j, \theta_j)\}_{j=1}^{\infty}$. Thus, according to the discrete representation of Q, it is appealing to assume that $\{(W_j, \theta_j)\}_{j=1}^{\infty}$ is driven by a two-parameter Poisson-Dirichlet process (Pitman and Yor, 1997), which can be written in stick-breaking form

(see, e.g. Ishwaran and James, 2001) as

$$W_1 = V_1, \quad \text{and} \quad W_j = V_j \prod_{l=1}^{j-1}(1 - V_l), \text{ for } j \geq 2,$$
$$\theta_j \overset{\text{i.i.d.}}{\sim} Q_0(\theta) \quad \text{and} \quad V_j \overset{\text{ind}}{\sim} \text{Be}(v_j | 1 - \alpha_{\text{PDP}}, j\alpha_{\text{PDP}}), \tag{5.6}$$

for $j \in \mathcal{J}$, where Q_0 (the baseline prior probability measure) is chosen to be a gamma distribution with density $q_0(\theta) \propto \theta^{a_{Q_0}-1} \exp\{-\theta b_{Q_0}\}$, for some fixed positive parameters a_{Q_0} and b_{Q_0}. The parameter α_{PDP} would also be fixed at some value in the interval $[0, 1)$.

5.3 Posterior Gibbs-sampler

Implementing a posterior sampling scheme for the model still requires an additional treatment to deal with the countable support of the latent index variables, (d_t), and the infinite sum involved in the denominator of the full conditional distribution. We overcome this problem by introducing further latent truncation variables. Details regarding the truncation scheme and derivation of full conditional distributions are given in Appendix C.

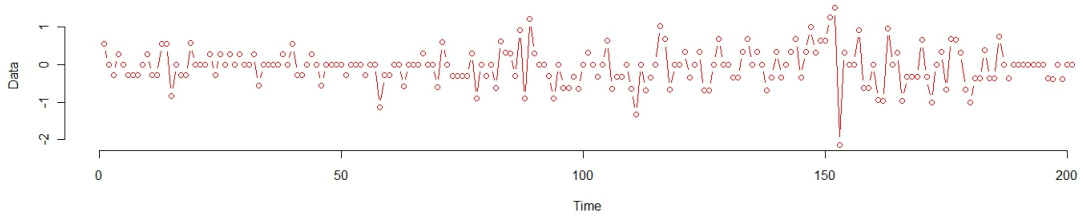
Posterior sampling can be carried out via the Gibbs sampler, with some trans-dimensional sampling steps. The sampler performs over the following full conditional distributions:

- $\boldsymbol{u}|\ldots$, which depends only on $\boldsymbol{d}$
- $\boldsymbol{d}|\ldots$, which is determined by particular full conditionals for $d_t|W_{d_t}, \theta_{d_t}, \lambda_t, \lambda_{t-1}, \mu, u_t, z_t$
- $\mu|\ldots$, which will involve a suitable trans-dimensional sampling scheme
- $Q|\ldots$, which follows the blocked Gibbs sampler scheme.
- $\boldsymbol{z}|\ldots$, which appear only in the cases where $d_t = d_{t-1}$
- $\boldsymbol{\lambda}|\ldots$, which will be case sensitive to different combinations induced by $(d'_{t-1}, d'_t, d'_{t+1})$

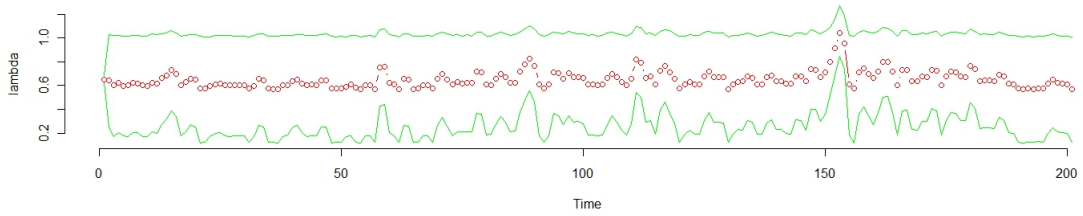
Full details regarding the implementation of the posterior sampler are given in Appendix C.

6 Empirical results

The stationary scale-mixture model is illustrated in practice with the analysis of a short segment of the Euro/USD weekly log-returns consisting of 200 observations[1]. The time-series data are displayed in Figure 1, in panel (a). No specificity about the period studied is relevant at this point, as our aim in this section is to show the performance of the model and the posterior sampling scheme. It is possible to see that the time-series data exhibit an initial period of stability followed by severe fluctuations towards the end of the observation period, which can be associated to periods with high volatility. All the results produced and reported bellow were obtained via the posterior Gibbs sampler with 15,000 iterations and considering a burn-in period of 10,000 iterations.



(a) Data



(b) Latent process $\{\lambda_t\}$

Figure 1: Euro/USD weekly log-returns and posterior estimate for the latent process (posterior means–dotted line, 95% confidence intervals–solid lines).

The model is parametrized by the random distribution Q, which is assumed to be driven by the two-parameter Poisson-Dirichlet process, and a $(0, 1)$-valued piecewise function μ,

[1]The data source is the European Central Bank, URL: `http://www.ecb.int`.

which is driven by an increasing in mean homogeneous marked Poisson process. Let us now elaborate on the intuition regarding the prior specification of the two-parameter Poisson-Dirichlet process, it can be seen that it belongs to the class of species sampling models[2]. As such, it has associated a posterior clustering scheme which shall be determined by the parameter $0 \leq \alpha_{\text{PDP}} < 1$, which can be used to tune the clustering scheme and give more probability to observe a new cluster as a new observation arrives. The larger the α_{PDP} is the more likely a new cluster will be identified. In our case we set this parameter at 0.5. The additional parameter corresponds to the baseline gamma distribution, Q_0, with shape and scale parameters given by a_{Q_0} and b_{Q_0}, respectively. Recall that draws of the baseline distribution determine the mixing distributions involved in the corresponding mixed-marginal density (3.1). We set $a_{Q_0} = 2$ and $b_{Q_0} = 3.3$ in order to reflect our prior information regarding the marginal dispersion of the data. Related to the posterior sampling scheme defined around Q is the sampling parameter ϕ, which basically serves to set the expected number of components in the mixture model that will be updated at each iteration. We set this parameter at $\phi = 0.18$, which will give the algorithm the possibility of exploring up to five more components at each iteration beyond the current state. The clustering scheme can be visualized trough the number and frequency of clusters summarized at each iteration through the vector of probability weights (W_j). In Figure 2 some of the MCMC sample paths for (W_j) are displayed. There we can observe that around ten clusters are being identified in the data at different iterations, with probability of membership determined by the magnitude of each W_j. There we can also observe that more clusters can be identified but with low probability. Those clusters have the possibility of catching future data and hence gain more weight, or loose it if the new data resembles more the data already observed.

With regards to the homogeneous Poisson marked process, its probability rule is characterized by the parameters ν_ξ, determining the intensity of the underlying Poisson process, and the triplet (a_μ, b_μ, c_μ), which determines the realization of the marked-part of the process. In order to get a reasonably large expected number of partitions in the domain of μ we set $\nu_\xi = 23$. On the other hand, the magnitude of the marked-part of the process, which

[2]Species sampling models are discrete random distribution functions whose random locations and random weights are mutually independent. See Pitman (1996) and Lijoi and Prünster (2010).

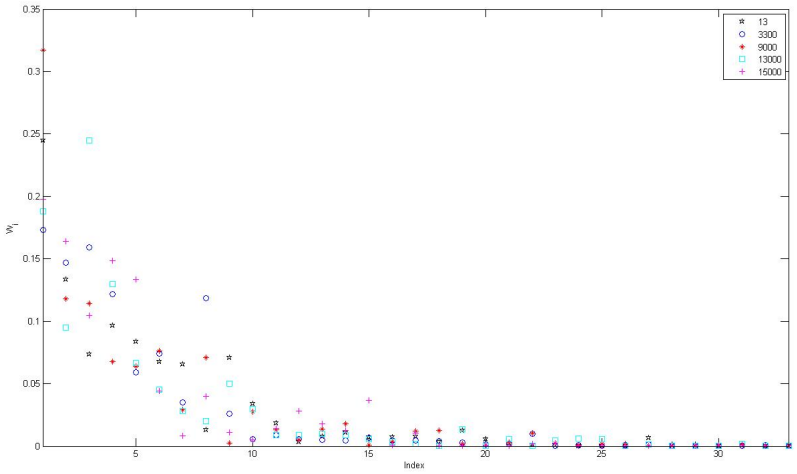


Figure 2: Some sample paths of (W_j) for the stationary scale-mixture model displayed at different iterations of the posterior sampling scheme.

measures the level of the random function μ, can be set at $a_\mu = 3.6$, $b_\mu = 1$ and $c_\mu = 4.5$. Recall that a_μ and $b_m u$ determine the distribution of the initial mark of the process. On the other hand, a_μ and c_μ determine the law of the evolution of the marked part. In particular, c_μ regulates how fast the function μ goes to 1 in mean[3]. Thus, with the above specification we guarantee that the sample paths of μ will go to 1 no so fast. The sampling procedure related to the trans-dimensional step involved in μ relies on q_{rj}, which basically regulates the probability of increasing or decreasing the number of components involved in μ in one unit with respect to the current state of the chain for this parameter. In order to give flexibility and balance to those moves we set $q_{rj} = 0.5$. Some sample paths derived from the posterior sampling scheme are displayed in Figure 3. There we can observe the way the sample paths of function μ adapt at each iteration. The plot summarized the posterior distribution for μ, and hence for G_1, which has low dispersion towards the region with high values for the latent variable λ.

Regarding the latent process (λ_t), it can be seen in Figure 1, panel (b), that the model captures the volatility features of the observed data. That is to say, the latent process reflects

[3]We impose the restriction on μ to go to 1 in mean in order to guarantee that the realized paths of the conditional random probability measure G_1 would satisfy regularity conditions of distribution functions.

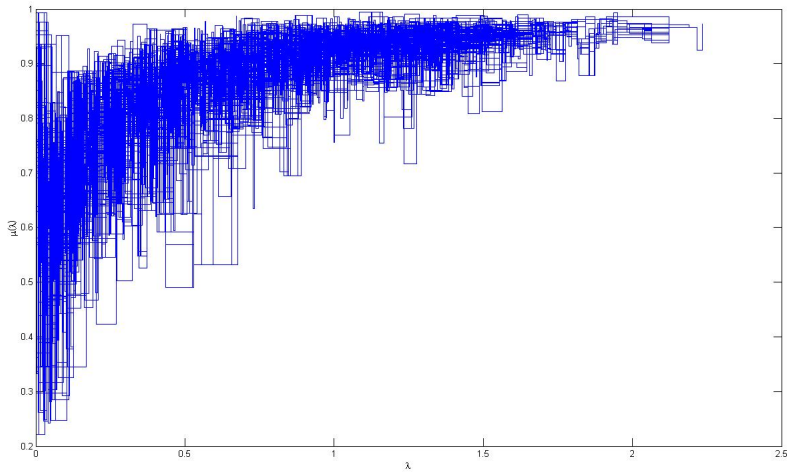
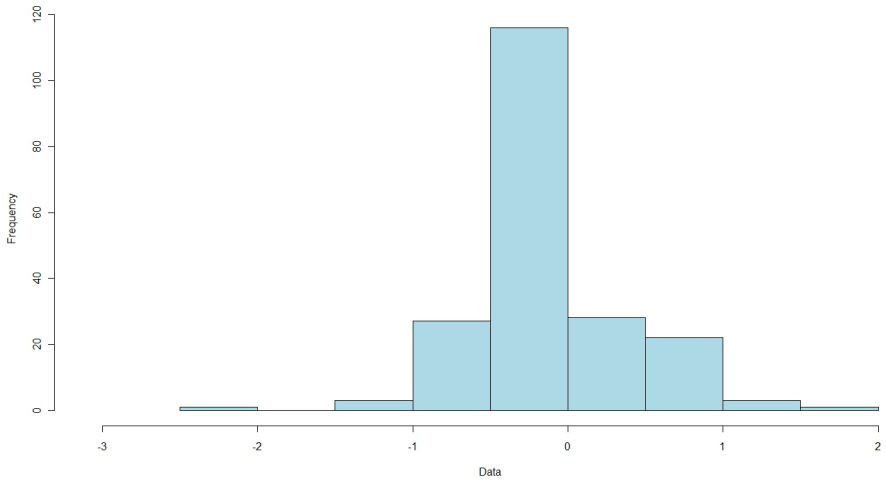


Figure 3: Sample paths of the function μ for the stationary scale-mixture model displayed at different iterations of the posterior sampling scheme.

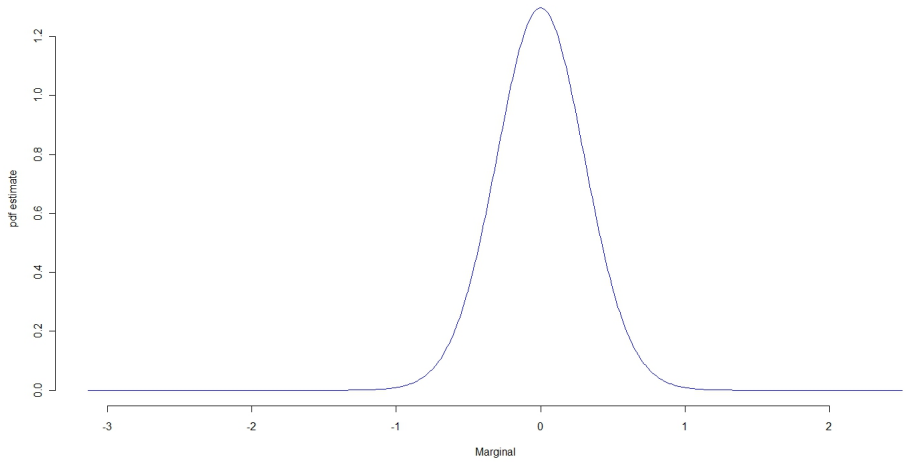
periods of high volatility and also periods with relative stability in the data. The same figure summarizes the sequence of posterior transition distributions $G_1(\lambda_t|\lambda_{t-1})$'s through the sequence of posterior means, i.e. $\mathbb{E}(\lambda_t|\lambda_{t-1}, \text{data})$ (dotted-lines in the plot) and their associated 95% posterior confidence bands (upper and lower solid-lines). Let us recall that G_1 is non-parametrically specified through Q and μ, hence no specific structural forms on G_1 are imposed, i.e. the distribution will adapt in the light of new information. Therefore, it is not surprising to observe asymetric confidence bands around the posterior means of the latent process.

In Figure 4, panel (a), we display the histogram of the 200 observations of the Euro/USD weekly log-returns. As it can be seen, the data show relatively heavy tails and it seems to be asymmetric. We display the histogram considering as if the observations were independent, in order to get some insights about their marginal behaviour, as it is the only empirical evidence available for this. However, that information is not entirely accurate, as the data are assumed to be dependent. On the other hand, it is shown in panel (b) of the same figure that the stationary scale-mixture captures part of the heavy tail behaviour of the data, however the asymmetry is not well reflected. We could expect that because by construction the marginal distribution of the model would be unimodal and symmetric, as the mixed-

kernel component is. It would be possible to provide the model with more structure in order to capture skewness in the data, but that is a matter of further work. The marginal predictive distribution displayed in panel (b) of the same figure was computed via Monte Carlo methods.



(a) Euro/USD weekly log-returns



(b) Posterior estimate of the marginal density

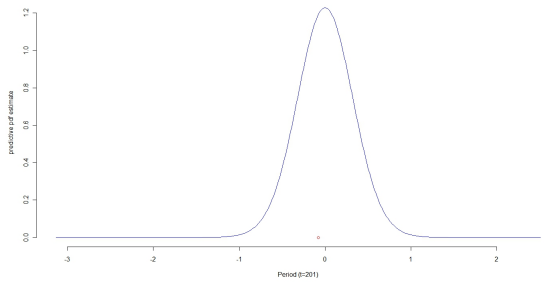
Figure 4: Histogram of the Euro/USD weekly log-returns and posterior estimate of their marginal density.

The predictive behaviour of the stationary scale-mixture model is summarized in a sequence of one-step predictive distributions for the period from $t = 201$ to $t = 206$. The

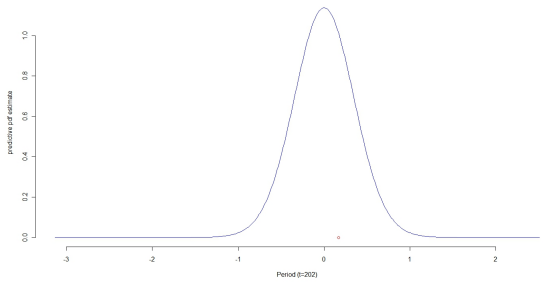
predictive distributions are computed as the posterior estimates for $p(y_{t+1}|y_t)$ given the data, i.e. they are defined as $\mathbb{E}\{p(y_{t+1}|y_t)|\boldsymbol{y}_t\}$ for $t = 200, \ldots, 205$ with $\boldsymbol{y}_t = (y_i : i = 1, \ldots, t)$. The Monte Carlo estimates for the predictive densities are displayed in Figure 5 along with the actual data for each t. Like the marginal, the predictive densities of the scale-mixture model are unimodal and symmetric, as the transitions dynamics are driving the spread or dispersion of the model conditionally on the previous observation. Once more, the model can be adapted in order to capture asymmetries or multi-modalities in the transition dynamics of the data, if required. For the moment, we just model the conditional dispersion in the data. As it can be seen, the predictive densities adapt their dispersion and tail behaviour according to the data observed, and they manage to reflect the uncertainty for future values of the data one-period ahead. Although, it is fair to say that the period we have predicted showed low volatility. Hence, it is shown that the stationary scale-mixture model performs reasonably well in terms of prediction. Such an assertion can be easily validated by computing predictive summaries as suggested in Gelfand et al. (1992). However, the predictive densities displayed in Figure 5 provide us with solid evidence for the period of study[4].

In this section we have shown that the stationary scale-mixture model can be implemented in practice and that it produces sensible results with actual data. However, some remarks about its implementation are in order. Bayesian nonparametric models, in general, are implemented using intensive and sophisticated computational algorithms, which require time to produce results. With the advent of new technology that problem can eventually be overcome. On the other hand, although the stationary scale-mixture model is a well defined fully nonparametric model it has the aim at capturing marginal and conditional dispersion in the data, as commented above. Although, it is relatively straightforward to adapt the model

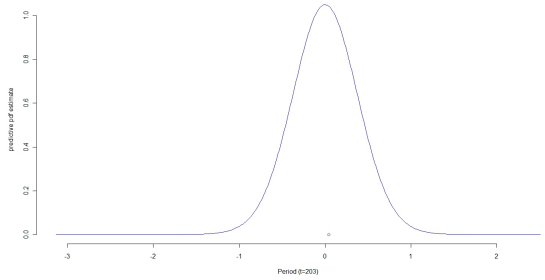
[4]Model assessment, within the Bayesian framework, can be done by means of comparing the actual values of future data, y_{t+1}'s, with their corresponding predictive distributions, $P(y_{t+1}|y_t, \text{data})$'s. Recall that $P(y_{t+1}|y_t, \text{data})$ comprises all one's updated beliefs about y_{t+1}. Alternatively, residual analysis (as traditionally used in practice) can be done by means of comparing the y_{t+1}'s with their expected outcomes, $\mathbb{E}(y_{t+1}|y_t)$'s (either predictive or adjusted), for instance. However, residual analysis may be restrictive as it makes reference just to a single característic of $P(y_{t+1}|y_t)$. Summary statistics for the latter comparison have been proposed by Gelman et al. (1996), among others; however no alternative summaries for Bayesian nonparametric models, in a general setting, are available so far.



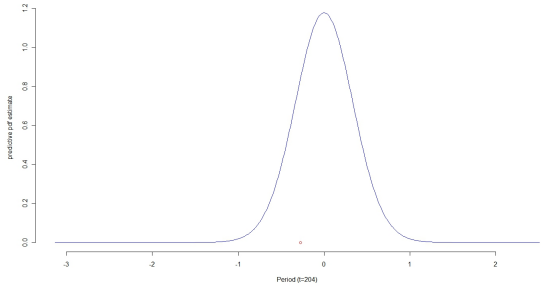
(a) t = 201



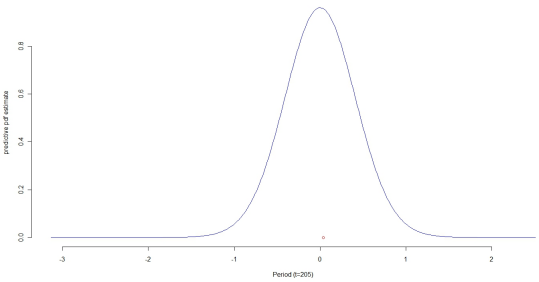
(b) t = 202



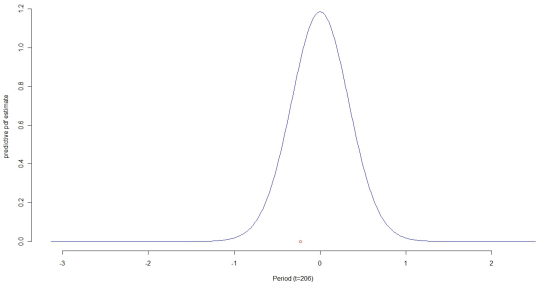
(c) t = 203



(d) t = 204



(e) t = 205



(f) t = 206

Figure 5: Sequence of one-step predictive densities for Euro/USD weekly log-returns from $t = 201$ to $t = 206$. The actual value of y_{t-1} for (a) is 0, for (b) is -0.0791, for (c) is 0.16983, for (d) is 0.04288, for (e) is -0.27747 and for (f) is 0.036403. The actual values of y_t are marked with "o" in the plots.

framework to different needs. That is the matter of our current and future work.

7 Discussion

In this paper we develop two general non-parametric strictly stationary time-series models. We give arguments that motivate the use of non-parametric stationary models in practice. However, the main contribution that this paper brings to the literature is the introduction of first-order stationary models admitting infinite-dimensional mixture representations for their marginal (invariant) and transition distributions. To the best of our knowledge, these are the first stationary models in the literature achieving that goal up to date.

Both models, benchmark and extended Gibbs sampler, can be adapted to a wide range of scenarios, as different specifications of their components can fulfill different needs. For example, different supports in observable and latent variables can be obtained by changing the specification of the mixed-kernel, in both models, and the baseline mixing distribution, in the case of the extended model, respectively. Regarding the extended model construction, different transition mechanisms can also be derived by means of considering alternative probability measures ruling the random distribution around which the dependence structure of the model is induced. The one induced by the beta-Stacy process is flexible enough, but it will be worth to explore other types of random probability measures driven by non-homogeneous Lévy process. Regarding the benchmark models, computational strategies to posterior inference are still under study.

The stationary scale-mixture model introduced in Section 4 corresponds to a particular case of the extended model construction. We developed an efficient algorithm for posterior inference. Predictions are obtained via traditional Monte Carlo methods (Robert and Casella, 1999). The marginal distribution that this model involves is unimodal and its tail behaviour is inherited from the tails of the base-line mixing distribution, in this case the Weibull distribution. The transition dynamics of this model can capture heavy-tails, due to the flexible specification of G_1. More general specifications aiming at capturing skewness in the data are possible by considering more structured parameterizations of the mixed-kernel component, for example.

In general, the models derived from the extended construction (the stationary scale-mixture model being a particular case) have an alternative state-space representation which resembles the well known hidden Markov models (see, e.g. Cappé et al., 2007), in particular the infinite-hidden Markov model (van Gael and Ghahramani, 2010). A key distinction of the models we develop in this paper with respect to traditional hidden Markov models is that the transition rules involved for observable and latent variables are regular conditional distributions. Hence, differently to traditional hidden-Markov models, in our models it is possible to recover the marginal dynamics for the observable process.

It is well known nowadays that it is required to establish conditions under which Bayesian nonparametric models, in general, attain posterior consistency (see, e.g. Diaconis and Freedman, 1986). It can be verified that the fully non-parametric stationary models introduced in this paper are Doob's type consistent (see, Lijoi et al., 2006). However conditions to obtain stronger consistency still need to be addressed, not just for our models but for most of the dependent Bayesian non-parametric models that have been developed so far. Studying posterior consistency is an open and active area in the field. The main challenge that our models have in this matter resides in the nonparametric representation of their marginal distributions, as traditional approaches to study large sample (asymptotic) properties are not applicable here. However, this problem is currently under study.

A number of generalizations of the models developed in this paper can be considered. For instance, it would be worth to explore multivariate generalizations, the incorporation of covariates, or modelling with higher order of dependencies. It would also be interesting to study the connection of these models with the continuous-time case.

References

Aldous, D. (1985). Exchangeability and related topics. In Hennequin, P. L., editor, *École d'Été de Probabilités de Saint-Flour XII*, Lecture Notes in Mathematics No. 117. Springer.

Arjas, E. and Gasbarra, D. (1994). Nonparametric Bayesian inference from right censored survival data using the Gibbs sampler. *Statistica Sinica*, 4:505–524.

Athreya, K. B. and Atuncar, G. S. (1998). Kernel estimation for real-valued Markov chains. *Sankhya: The Indian Journal of Statistics*, 60(A):1–17.

Barber, D., Cemgil, A. T., and Chiappa, S. (2008). *Proceedings of the Workshop Inference and Estimation in Probabilistic Time-Series Models*. Isaac Newton Institute for Mathematical Sciences, Cambridge.

Barber, D., Cemgil, A. T., and Chiappa, S. (2011). *Bayesian Time Series Models*. Cambridge University Press, Cambridge.

Barry, C. B. and Winkler, R. L. (1976). Nonstationarity and portfolio choice. *Journal of Financial and Quantitative Analysis*, 11:217–235.

Bouzar, N. (2010). Autoregressive sequences via Lévy processes. *REVSTAT - Statistical Journal*, 8(2):81–103.

Bouzar, N. and Jayakumar, K. (2008). Time series with discrete semistable marginals. *Statistical Papers*, 49(4):619–635.

Campos, V. S. M. and Dorea, C. C. Y. (2005). Kernel estimation for stationary density of Markov chains with general state space. *Annals of the Institute of Statistical Mathematics*, 57(3):443–453.

Cappé, O., Moulines, E., and Rydën, T. (2007). *Inference in Hidden Markov Models*. Springer, New York.

Chu, C.-S. J., Stinchocombe, M., and White, H. (1996). Monitoring structural change. *Econometrica*, 64(5):1045–1065.

Clements, M. P. and Hendry, D. F. (1999). *Forecasting Non-Stationary Economic Time Series*. The MIT Press, Cambridge, Mass.

Clements, M. P. and Hendry, D. F. (2008). Economic forecasting in a changing world. *Capitalism and Society*, 3(2):Article 1.

Contreras-Cristán, A., Mena, R. H., and Walker, S. G. (2009). On the construction of stationary AR(1) models via random distributions. *Statistics*, 43(3):227–240.

de Finetti, B. (1937). La previsión: Ses logiques, ses sources subjectives. *Annals of the Institute Henri Poincaré*, 7:1–68.

Diaconis, P. W. and Freedman, D. (1986). On the consistency of bayes estimates (with discussion). *The Annals of Statistics*, 14(1):1–67.

Doksum, K. (1974). Tailfree and neutral-to-the-right random probabilities and their posterior distributions. *The Annals of Probability*, 2(2):183–201.

Engle, R. F. (2001). GARCH 101: The use of ARCH/GARCH models in applied econometrics. *Journal of Economic Perspectives*, 15(4):157–168.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer, New York.

Gaver, D. P. and Lewis, P. A. W. (1980). First-order autoregressive gamma sequences and point processes. *Advances in Applied Probability*, 12(3):727–745.

Gelfand, A. E., Dey, D. K., and Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods. In Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M., editors, *Bayesian Statistics 5*, pages 147–167. Oxford University Press, Oxford.

Gelman, A., Meng, X. L., and Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Statistica Sinica*, 6:733–807.

Giacominia, R., Gottschling, A., Haefke, C., and White, H. (2008). Mixtures of t-distributions for finance and forecasting. *Journal of Econometrics*, 144:175–192.

Godsill, S. T. (2001). On the relationship between MCMC model uncertainty methods. *Journal of Computational and Graphical Statistics*, 10(2):230–248.

Haas, M., Mittnik, S., and Paolella, M. S. (2004). A new approach to Markov-switching GARCH models. *Journal of Financial Econometrics*, 2(4):493–530.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384.

Hili, O. (2001). Estimation non paramétrique de la densité d'un processus stationaire mélangeant. *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics*, 332(1):841–844.

Ishwaran, H. and James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 96(453):161–173.

Jacobs, P. A. and Lewis, P. A. (1978). Discrete time series generated by mixtures. I: Correlational and runs properties. *Journal of the Royal Statistical Society, Series B*, 40(1):94–105.

Joe, H. (1996). Time series with univariate margins in the convolution-closed infinitely divisible class. *Journal of Applied Probability*, 33(3):664–677.

Jørgensen, B. and Song, P. X.-K. (1998). Stationary time series models with exponential dispersion model margins. *Journal of Applied Probability*, 35(1):78–92.

Jørgensen, B. and Song, P. X.-K. (2006). Binomial thinning models for integer time series. *Statistical Modelling*, 6:81–96.

Krolzig, M.-H. (1997). *Markov Switching Vector Autoregressions: Modelling, Statistical Inference and Application to Business Cycle Analysis*. Springer, Berlin.

Lacour, C. (2008). Nonparametric estimation of the stationary density and the transition density of a Markov chain. *Stochastic Processes and their Applications*, 118(2):232–260.

Lijoi, A. and Prünster, I. (2010). Models beyond the Dirichlet process. In Hjort, N. L., Holmes, C. C., Müller, P., and Walker, S. G., editors, *Bayesian Nonparametrics*, pages 80–130. Cambridge University Press, Cambridge.

Lijoi, A., Prünster, I., and Walker, S. G. (2006). Bayesian consistency for stationary models. *Econometric Theory*, 23:479–759.

Lo, A. Y. (1984). On a class of Bayesian nonparametric estimates: I. Density estimates. *The Annals of Statistics*, 12(1):351–357.

Maitra, A. (1977). Integral representations of invariant measures. *Transactions of the American Mathematical Society*, 229:209–225.

Martínez-Ovando, J. C. (2011). *Contributions to the Bayesian Nonparametric Modelling of Time-Series Data*. PhD thesis, School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, United Kingdom.

Mena, R. H. and Walker, S. G. (2005). Stationary autoregressive models via a Bayesian nonparametric approach. *Journal of Time Series Analysis*, 26(6):789–805.

Mena, R. H. and Walker, S. G. (2007). On the stationary version of the generalized hyperbolic ARCH model. *Annals of the Institute of Statistical Mathematics*, 59:325–348.

Neal, R. M. (2003). Slice sampling (with discussion). *The Annals of Statistics*, 31(3):705–767.

Nieto-Barajas, L. E. and Walker, S. G. (2002). Markov beta and gamma processes for modelling hazard rates. *Scandinavian Journal of Statistics*, 29:413–424.

Pitman, J. (1996). Some developments of the Blackwell-MacQueen urn scheme. In Ferguson, T. S., Shapley, L. S., and MacQueen, J. B., editors, *Statistics, Probability and Game Theory*, pages 245–267. Institute of Mathematics and Statistics, Hayward, CA.

Pitman, J. and Yor, M. (1997). The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Annals of Probability*, 25(2):855–900.

Pitt, M. K., Chatfield, C., and Walker, S. G. (2002). Constructing first order stationary autoregressive models via latent processes. *Scandinavian Journal of Statistics*, 29:657–663.

Pitt, M. K. and Walker, S. G. (2005). Constructing stationary time series models using auxiliary variables with applications. *Journal of the American Statistical Association*, 100(470):554–564.

Pitt, M. K. and Walker, S. G. (2006). Extended construction of stationary autoregressive process. *Statistics & Probability Letters*, 76(12):1219–1224.

Robert, C. P. and Casella, G. (1999). *Monte Carlo Statistical Methods*. Springer-Verlag, London.

Rodríguez, A., Dunson, D. B., and Gelfand, A. E. (2008). The nested Dirichlet process. *Journal of the American Statistical Association*, 103(483):1131–1154.

Roussas, G. C. (1969). Nonparametric estimation of the transition distribution function of a Markov process. *The Annals of Mathematical Statistics*, 40(4):1386–1400.

Teh, Y. W., Jordan, M. I., Beal, M. J., and Blei, D. M. (2006). Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 101(476):1566–1581.

van Gael, J. and Ghahramani, Z. (2010). Nonparametric hidden Markov models. In Barber, D., Cemgil, A. T., and Chiappa, S., editors, *Bayesian Time-Series Models*. Cambridge University Press, Cambridge.

Walker, S. G. (2007). Sampling the Dirichlet mixture model with slices. *Communications in Statistics*, 36:45–54.

Walker, S. G. and Muliere, P. (1997). Beta-Stacy processes and a generalization of the Pólya-urn scheme. *The Annals of Statistics*, 25(4):1762–1780.

Zabell, S. L. (2005). *Symmetry and Its Discontents: Essays on the History of Inductive Probability*. Cambridge Studies in Probability, Induction, and Decision Theory. Cambridge University Press, Cambridge.

Zeileis, A., Leisch, F., Kleiber, C., and Hornik, K. (2005). An efficient method for estimating seemingly unrelated nonlinear regressions and tests for aggregation bias. *Journal of Applied Econometrics*, 20:99–121.

Zhu, R. and Joe, H. (2006). Modelling count data time series with Markov processes based on binomial thinning. *Journal of Time Series Analysis*, 27(5):725–738.

A Gibbs sampler model construction based upon mixtures

The model derived in Section 3 can be addressed in a more general framework based upon mixtures. It is possible to depart from a joint probability measure derived from the following

random density (or distribution) function

$$f_H(y) = \int K(y|\theta)H(\mathrm{d}\theta), \tag{A.1}$$

where θ is a generic mixing random variable, $K(y|\theta)$ is a fixed parametric kernel, and H is a random distribution function. Accordingly, the probability measure for f_H would be characterized by a probability measure for H. Hence, the analogous stochastic set-up to (2.5) would be given by

$$y|f_H \sim f_H, \quad \text{and} \quad H \sim \Pi, \tag{A.2}$$

where Π is a probability measure for H, with the additional assumption that $\mathbb{E}_\Pi(H) = G_0$, for a given diffuse distribution function G_0. After applying the Gibbs sampler model construction approach, marginal and first-order transition distributions of a stationary model would be

$$p(y_t) = \int f_H(y_t)\Pi(\mathrm{d}H) = \int K(y_t|\theta)G_0(\mathrm{d}\theta), \tag{A.3}$$

$$p(y_t|y_{t-1}) = \int f_H(y_t)\Pi(\mathrm{d}H|y_{t-1}) = \int K(y_t|\theta)G_1(\mathrm{d}\theta|\theta')G_0(\mathrm{d}\theta'|y_{t-1}), \tag{A.4}$$

where θ and θ' are two auxiliary mixing variables, $G_1(\mathrm{d}\theta|\theta') = \mathbb{E}_\Pi(H(\mathrm{d}\theta)|\theta')$ and $G_0(\theta'|y_1) \propto K(y_1|\theta')G_0(\theta')$. As a consequence, the pair (A.3) and (A.4) characterize a strictly stationary Markov process.

A key feature of the stationary models derived upon mixtures is that it is actually possible to express their transition dynamics in terms of observable and latent variables together by extending the scope to (A.2) to the augmented probability measure induced by

$$y|\theta \sim K(y|\theta), \quad \theta|H \sim H, \quad H \sim \Pi. \tag{A.5}$$

Therefore, joint marginal (invariant) and transition distributions for the duet (Y, θ) can be written as

$$p(y_t, \theta_t) = \int K(y_t|\theta_t)H(\theta_t)\Pi(\mathrm{d}H) = K(y_t|\theta_t)G_0(\theta_t), \tag{A.6}$$

$$p(y_t, \theta_t|y_{t-1}, \theta_{t-1}) = \int K(y_t|\theta_t)H(\theta_t)\Pi(\mathrm{d}H|\theta_{t-1}) = K(y_t|\theta_t)G_1(\theta_t|\theta_{t-1}). \tag{A.7}$$

The latter relation follows from the conditional independence structure between Y and H, given θ. It is also verifiable that (A.6)-(A.7) characterize the probability law of the joint

Markov sequence $\{(Y_t, \theta_t)\}_{t=1}^{\infty}$. This construction resembles the specification of state-space models. However, in our context, the marginal part of the joint model corresponding to the observables , i.e. $\{Y_t\}_{t=1}^{\infty}$, can be recovered from (A.7) and coincides with (A.4), due that the conditional distributions involved in our model construction are regular.

Finally, it can be seeing that the extended model construction we developed in Section 3 was raised from a more structured version to (A.5). We now proceed with a detailed description of the formulae involved in the stationary scale-mixture model presented in Section 4.

B Stationary scale mixture model (relevant formulae)

In this section we provide with relevant formulae regarding the stationary scale-mixture model developed in Section 4. We give explicit analytical expressions for the transition dynamics of the model and develop on further details regarding in the stochastic truncation scheme.

B.1 Transition dynamics

The transition dynamics of the stationary scale-mixture model are characterized by the distribution given in (4.8), which corresponds to a mixed-type transition whose density/probability-mass function admit the following decomposition

$$g_1(\lambda|\lambda') = \sum_{z=-1}^{1} g_{1,z}(\lambda|\lambda')p(z|\lambda'), \tag{B.1}$$

where λ and λ' are two latent variables and z is a further latent variable which indicates the region where λ takes values after considering the partition induced on its support by conditioning on λ'. The other components involved in the mixture are

$$\begin{aligned}
p(z=-1|\lambda') &= 1 - S_\mu(\lambda'), & g_{1,z=-1}(\lambda|\lambda') &= \frac{g_\mu(\lambda|\lambda')}{1-S_\mu(\lambda')}\,\mathbb{1}(\lambda<\lambda'),\\
p(z=0|\lambda') &= [1-\mu(\lambda')]\,S_\mu(\lambda'), & g_{1,z=0}(\lambda|\lambda') &= \delta_{\lambda'}(\lambda),\\
p(z=1|\lambda') &= \mu(\lambda')S_\mu(\lambda'), & g_{1,z=1}(\lambda|\lambda') &= \frac{g_0(\lambda)}{S_0(\lambda')}\,\mathbb{1}(\lambda>\lambda');
\end{aligned} \tag{B.2}$$

with,

$$S_\mu(\lambda) = \exp\left\{-\sum_{\{k:\xi_k\leq\lambda'\}}\mu_k h_{0,k}(\lambda)\right\} \tag{B.3}$$

$$h_{0,k}(\lambda) = \int_{E_k\cap(0,\lambda]} h_0(s)\mathrm{d}s, \tag{B.4}$$

$$S_0(\lambda) = \exp\left\{-\int_0^\lambda h_0(u)\mathrm{d}u\right\}, \tag{B.5}$$

and

$$g_\mu(\lambda|\lambda') = \mu(\lambda)h_0(\lambda)S_\mu(\lambda)\,\mathbb{1}(\lambda<\lambda') \tag{B.6}$$

$$= \mu_{k(\lambda)}h_0(\lambda)\cdot\exp\left\{-\sum_{\{k:\xi_k\leq\lambda'\}}\mu_k h_{0,k}(\lambda)\right\}, \tag{B.7}$$

$$g_0(\lambda) = h_0(\lambda)S_0(\lambda). \tag{B.8}$$

The index $k(\lambda)$ involved in (B.7) stands for the index of the partition associated with μ at which $\lambda\in E_k$, and $\mu_{k(\lambda)}$ stands for its associated level, with the $h_{0,k}(\lambda)$ given. The function h_0 refers to the hazard function of the baseline distribution G_0 –i.e. the Weibull distribution in this case–, and g_0 its associated density.

B.2 Stochastic truncation via latent variables

The stochastic truncation scheme seems to be a sensible way of dealing with the infinite-sums involved in the likelihood (4.10). The truncation is carried out through (5.2), where the latent

variable U_t is supported in the set $\{0, 1, \ldots\}$. Accordingly, the augmented likelihood become

$$
\begin{aligned}
lik \;\propto\; & \left\{\prod_{t=1}^{T} \mathrm{N}(y_t|0,\lambda_t)\right\} \cdot \left\{\prod_{\{t=1\}\cup\{t:d_t\neq d_{t-1}\}} W_{d_t}(1-\phi)^{u_t-d_t}\,\mathbb{1}(u_t\geq d_t)\theta_{d_t}\lambda_t^{\theta_{d_t}-1}e^{-\lambda_t^{\theta_{d_t}}}\right\} \\
& \cdot \left\{\prod_{\{t:d_t=d_{t-1},z_t=-1\}} W_{d_t}(1-\phi)^{u_t-d_t}\,\mathbb{1}(u_t\geq d_t)\mu_{k(\lambda_t)}\theta_{d_t}\lambda_t^{\theta_{d_t}-1}\right\} \\
& \cdot \left\{\prod_{\{t:d_t=d_{t-1},z_t=0\}} W_{d_t}(1-\phi)^{u_t-d_t}\,\mathbb{1}(u_t\geq d_t)(1-\mu_{k(\lambda_{t-1})})\right\} \\
& \cdot \left\{\prod_{\{t:d_t=d_{t-1},z_t=1\}} W_{d_t}(1-\phi)^{u_t-d_t}\,\mathbb{1}(u_t\geq d_t)\mu_{k(\lambda_{t-1})}\theta_{d_t}\lambda_t^{\theta_{d_t}-1}e^{-(\lambda_t^{\theta_{d_t}}-\lambda_{t-1}^{\theta_{d_t}})}\right\} \\
& \cdot \exp\left\{-\sum_{\{k:\xi_k<\lambda_t\}}\sum_{\{t:d_t=d_{t-1},z_t=-1\}} \mu_k h_{0,k}(\lambda_t)\right\} \\
& \cdot \exp\left\{-\sum_{\{k:\xi_k<\lambda_{t-1}\}}\sum_{\{t:d_t=d_{t-1},z_t\in\{0,1\}\}} \mu_k h_{0,k}(\lambda_{t-1})\right\}, \qquad \text{(B.9)}
\end{aligned}
$$

which is defined for $(Q, \mu, \boldsymbol{\lambda}, \boldsymbol{z}, \boldsymbol{d}, \boldsymbol{u})$, where $\boldsymbol{u} = \{u_t\}_{t=1}^{T}$ stands for the sequence of latent truncation variables, and $k(\lambda_t)$ and $h_{0,k}(\lambda_t)$ are given. Further details about this and other stochastic truncation schemes involving infinite mixture models can be found in Martínez-Ovando (2011). In the next appendix we will elaborate on the algorithm for posterior sampling developed around (B.9) for the stationary scale-mixture model.

C Further details of the posterior sampler

Here we provide with details regarding the full conditional distributions and posterior sampling scheme sketched in Section 5.3, related to the stationary scale-mixture model.

C.1 Full conditional distribution for $\boldsymbol{u}$

The full conditional distribution for $\boldsymbol{u}$ depends only upon $\boldsymbol{d}$. Moreover, the u_t's are conditionally independent given the indexes d_t's. Therefore, we only need to specify the generic full conditional distribution for each u_t which, conditionally on d_t, follows the shifted geometric

distribution

$$\pi(u_t|d_t) = \phi(1-\phi)^{u_t-d_t}, \tag{C.1}$$

with support in the set $\{d_t, d_t+1, \ldots\}$, as given in (5.2). Sampling from this full conditional distribution is doable using standard procedures.

C.2 Full conditional distribution for $\boldsymbol{d}$

The full conditional distribution for $\boldsymbol{d}$ is defined by the individual full conditionals for each d_t, which are derived around (5.3). But, differently to standard nonparametric Bayesian mixture models, the conditional independence structure among the d_t's does not prevail in our model. Instead, these latent indexes exhibit a Markovian dependence structure. Therefore, the full conditional distribution for each d_t will depend on $(W_{d_t}, \theta_{d_t}, \lambda_t, \lambda_{t-1}, \mu, u_t)$ and the pair of indexes (d_{t-1}, d_{t+1}). Hence, d_t would be distributed by

$$\pi(d_t|\ldots) \propto g(d_t)(1-\phi)^{-d_t}\,\mathbb{1}(d_t \leq u_t), \tag{C.2}$$

for each t, where $g(d_t)$ is a measurable function which represents the general component of the likelihood in which d_t is involved. [Notice that in general mixture models the function $g(d_t)$ would be given by $W_{d_t} f(\lambda_t|\theta_{d_t})$, where $f(\lambda_t|\theta_{d_t})$ would be the kernel component of the mixture. But that is not our case.]

Before proceeding, notice that conditioning on u_t truncates the range of possible values that d_t can take to the set $\{1, \ldots, u_t\}$. Hence, the full conditional distribution for d_t will involve calculating finite-sums of the form

$$\sum_{d_t=1}^{u_t} g(d_t)(1-\phi)^{-d_t}.$$

The factor $g(d_t)$ corresponds to the product of mixed-kernel and random weight associated with the d_t-th component. As mentioned before, the contribution of d_t's to the likelihood would lead us to functions $g(d_t)$ defined as

$$g(d_t = j) = \begin{cases} W_j g_1(\lambda_{t+1}|\lambda_t;\theta_j,\mu) g_1(\lambda_t|\lambda_{t-1};\theta_j,\mu) & , \quad j = d_{t+1} = d_t = d_{t-1}, \\ W_j g_1(\lambda_{t+1}|\lambda_t;\theta_j,\mu), & , \quad j = d_{t+1} = d_t,\ d_t \neq d_{t-1}, \\ W_j g_1(\lambda_t|\lambda_{t-1};\theta_j,\mu), & , \quad d_{t+1} \neq d_t,\ j = d_t = d_{t-1}, \\ W_j\, \mathrm{We}(\lambda_t|\theta_j), & , \quad j = d_t,\ d_{t+1} \neq d_t \neq d_{t-1}, \end{cases} \tag{C.3}$$

for $j = 1, 2, \ldots, u_t$. Therefore, sampling a new d_t, conditionally on u_t, will involve drawing samples from a discrete distribution with support at $\{1, ..., u_t\}$ and probability weights proportional to the factors $g(d)(1-\phi)^{-d}$, fot different values of d.

Two obvious exceptions to the specification of $g(d_t)$ given above are in order for the times $t = 1$ and T. For the initial time, $t = 1$, this function will take the form

$$g(d_1 = j) = W_j \, \mathrm{We}(\lambda_1 | \theta_j),$$

for $j = 1, 2, \ldots, u_1$, as it happens with traditional mixture models. As for the time $t = T$, the function g becomes

$$g(d_T = j) = \begin{cases} W_j g_1(\lambda_T | \lambda_{T-1}; \theta_j, \mu), & , \quad j = d_T = d_{T-1}, \\ W_j \, \mathrm{We}(\lambda_T | \theta_j), & , \quad j = d_T, \ d_T \neq d_{T-1}, \end{cases}$$

for $j = 1, 2, \ldots, u_T$. Notice that the above distribution would be specified in terms of the different combinations that the pair (z_t, z_{t+1}) may also generate. For the cases where either $d_t = d_{t-1}$ or $d_t = d_{t+1}$, such a distribution will be given in terms of z_t or z_{t+1}, respectively.

C.3 Full conditional distribution for μ

The full conditional distribution for μ does not have a closed analytic form. Updating μ will require the implementation of trans-dimensional samplers. The one we adopt here updates the sequence $\{(\xi_k, \mu_k)\}$ according to the current latent variables $\{\lambda_t\}$ corresponding to the cases $\{t : d_t = d_{t-1}\}$. At each iteration, the form of the full conditional distribution for μ will be given by

$$\begin{aligned} \pi\left(\{(\xi_k, \mu_k)\}_{k=0}^{\infty} | (\lambda_t, z_t, \theta_{d_t}, d_t : t = 1 \ldots, T)\right) \quad \propto \quad & lik\left(\{(\xi_k, \mu_k)\}_{k=0}^{m}; (\lambda_t, z_t, \theta_{d_t}, d_t : t = 1 \ldots, T)\right) \\ & \cdot \pi\left(\{(\xi_k, \mu_k)\}_{k=0}^{m}\right) \pi(m | \lambda^*) \end{aligned} \tag{C.4}$$

where $m = \#\{\xi_k \leq \lambda^*\}$, with $\lambda^* = \max_t\{\lambda_t\}$. Regarding the prior specification for μ given in Section 5.2.1, after conditioning on the data will induce the alternative representation

$$\begin{aligned} \pi(m | \lambda^*) \quad & \propto \quad \nu_\xi^m e^{-\nu_\xi \lambda^*}, \\ \pi\left(\{(\xi_k, \mu_k)\}_{k=0}^{m} | m\right) \quad & \propto \quad \mathrm{Be}(\mu_0 | a_\mu, b_\mu) \cdot \left\{ \prod_{k=1}^{m} \mathrm{Be}\left(\mu_k | a_\mu + k, c_\mu\right) \right\}, \end{aligned} \tag{C.5}$$

where m is random. Hence, the part of the likelihood (C.4) involving μ becomes

$$
\begin{aligned}
lik(\{(\xi_k,\mu_k)\}_{k=0}^{m};(\lambda_t,z_t,\theta_{d_t},d_t:t=1\ldots,T)) \propto \\
\propto \left\{\prod_{\{t:d_t=d_{t-1},z_t\in\{-1,1\}\}} \mu_{k(\lambda_t)}\right\}\cdot\left\{\prod_{\{t:d_t=d_{t-1},z_{-t}=0\}}(1-\mu_{k(\lambda_{t-1})})\right\} \\
\cdot\exp\left\{-\sum_{\{k:\xi_k<\lambda_t\}}\sum_{\{t:d_t=d_{t-1},z_t=-1\}}\mu_k h_{0,k}(\lambda_t)\right\} \\
\cdot\exp\left\{-\sum_{\{k:\xi_k<\lambda_{t-1}\}}\sum_{\{t:d_t=d_{t-1},z_t\in\{0,1\}\}}\mu_k h_{0,k}(\lambda_{t-1})\right\},
\end{aligned}
\tag{C.6}
$$

where $k(\lambda_t)$ and $h_{0,k}(\lambda_t)$ are given as before, for $k=1,2,\ldots$.

The trans-dimensional sampler we implement follows ideas developed by Godsill (2001), which consist on setting a Gibbs-Metropolis sampler over an infinite-dimensional parameter space. That scheme requires to extend the scope to (C.6) by including a further latent variable, v, indicating the component of $\{(\xi_k,\mu_k)\}_{k=0}^{\infty}$ at which the sampler shall move. Conditionally on the current number of components m involved in μ and on $\boldsymbol{\lambda}$. The distribution for v will be given by

$$
v|m=\begin{cases} m, & \text{with prob. } 1-q_{rj}, \\ m+1, & \text{with prob. } q_{rj}, \end{cases}
\tag{C.7}
$$

where $0<q_{rj}<1$ is a pre-set sampling parameter.

Therefore, the updating scheme for μ will consist in moving the current state of the chain $\{(\xi_k,\mu_k)\}_{k=0}^{m}$ to $\{\xi_k',\mu_k'\}_{k=0}^{'}$, with m' being defined on $\{m-1,m,m+1\}$. This step is restricted to the interval $(0,\lambda^*]$, which varies across iterations. The updating procedure involves the following steps:

i) Conditionally on m, proposed states $\{(\xi_k',\mu_k')\}_{k=1}^{l}$ are drawn for $l=m-1,m,m+1$. This step is described bellow.

ii) Conditionally on m, the parameter v' is updated from (C.7).

iii) Conditionally on v', the index m' is updated by drawing a sample from its full conditional distribution, which is defined by the updated value of v', that is: a) $v'=m$, and b) $v'=m+1$. Those distributions are given below.

a) In the case where $u' = m$, the full conditional distribution for m' becomes

$$m'|(v' = m) = \begin{cases} m-1, & \text{with prob.} \quad \propto q_{rj} p\left(\{(\xi'_k, \mu'_k)\}_{k=0}^{m-1}|\ldots\right) \\ & \qquad \cdot p\left(\{(\xi'_k, \mu'_k)\}_{k=0}^{m}|\{(\xi'_k, \mu'_k)\}_{k=0}^{m-1}, m\right), \\ m, & \text{with prob.} \quad \propto (1-q_{rj}) p\left(\{(\xi'_k, \mu'_k)\}_{k=0}^{m}|\ldots\right) \\ & \qquad \cdot p\left(\{(\xi'_k, \mu'_k)\}_{k=0}^{m-1}|\{(\xi'_k, \mu'_k)\}_{k=0}^{m}, m\right). \end{cases} \tag{C.8}$$

b) The case $v' = m+1$ gives rise to the following conditional distribution for m',

$$m'|(v' = m+1) = \begin{cases} m, & \text{with prob.} \quad \propto q_{rj} p\left(\{(\xi'_k, \mu'_k)\}_{k=0}^{m}|\ldots\right) \\ & \qquad \cdot p\left(\{(\xi'_k, \mu'_k)\}_{k=0}^{m+1}|\{(\xi'_k, \mu'_k)\}_{k=0}^{m}, m\right), \\ m+1, & \text{with prob.} \quad \propto (1-q_{rj}) p\left(\{(\xi'_k, \mu'_k)\}_{k=0}^{m+1}|\ldots\right) \\ & \qquad \cdot p\left(\{(\xi'_k, \mu'_k)\}_{k=0}^{m}|\{(\xi'_k, \mu'_k)\}_{k=0}^{m+1}, m\right). \end{cases} \tag{C.9}$$

iv) Given the current value of m',proposed states $\{(\xi'_k, \mu'_k)\}_{k=0}^{m'}$ are drawn and kept as the updated state of m'.

In the above expressions, $p\left(\{(\xi'_k, \mu'_k)\}_{k=0}^{l}|\ldots\right) = p\left(\{(\xi'_k, \mu'_k)\}_{k=0}^{l}|(\lambda'_t, z'_t, \theta'_{d_t}, d'_t : t = 1\ldots, T)\right)$, for $l = m-1, m, m+1$.

Concerning the step (i) described above, three proposed moves are defined in order to draw new states $\{(\xi'_k, \mu'_k)\}_{k=0}^{l}$ for $l = m-1, m, m+1$: i) Updating a current change point; ii) incorporating a new change point; and iii) deleting a change point. Details are given below.

C.3.1 Updating a current change point.

Proposing $\{(\xi'_k, \mu'_k)\}_{k=0}^{m}$, given m, consists in choosing an index k' at random from the set $\{0, 1, \ldots, m\}$. Then, a new level μ'_k is drawn from the distribution

$$p(\mu'_{k'}|\mu_{k'-1}, \mu_{k'+1}, \xi_{k'}) = \mathrm{U}(\mu'_{k'}|\min\{\mu_{k'-1}, \mu_{k'+1}\}, \max\{\mu_{k'-1}, \mu_{k'+1}\}), \tag{C.10}$$

and the current index $\xi'_k = \xi_{k'}$ is kept. The other components are updated as $\xi'_k = \xi_k$ and $\mu'_k = \mu_k$, for $k \neq k'$. Then, the updated sequence $\{(\xi'_k, \mu'_k)\}_{k=0}^{m}$ is accepted with probability

$$\min\left\{1, \frac{p\left(\boldsymbol{\lambda}|\{(\xi'_k, \mu'_k)\}_{k=0}^{m}; \boldsymbol{z}, \theta\right) \pi\left(\{(\xi'_k, \mu'_k)\}_{k=0}^{m}|m\right)}{p\left(\boldsymbol{\lambda}|\{(\xi_k, \mu_k)\}_{k=0}^{m}; \boldsymbol{z}, \theta\right) \pi\left(\{(\xi_k, \mu_k)\}_{k=0}^{m}|m\right)}\right\}, \tag{C.11}$$

where $p\left(\boldsymbol{\lambda}|\{(\xi_k,\mu_k)\}_{k=0}^m,\boldsymbol{z},\theta\right)$ and $\pi\left(\{(\xi_k,\mu_k)\}_{k=0}^m|m\right)$ are given as in (C.6) and (C.5), respectively.

C.3.2 Incorporating a new change point.

Drawing $\{(\xi'_k,\mu'_k)\}_{k=0}^{m+1}$, given m, consists in sampling a new change point ξ' from the distribution $\mathrm{U}(0,\lambda^*)$ and, given ξ', a new index $k(\xi')=\{k:\xi'\in E_k, k=1,\ldots,m\}$ is defined. Then, we propose a new level μ' associated with ξ' by drawing a sample from

$$p(\mu'|\xi',\mu_{k(\xi')},\mu_{k(\xi')+1}) \;=\; \mathrm{U}(\mu'_{k'}|\min\{\mu_{k'-1},\mu_{k'+1}\},\max\{\mu_{k'-1},\mu_{k'+1}\}). \qquad \text{(C.12)}$$

Let us assume that $k(\xi')=l$. Then, the proposed sequence $\{(\xi'_k,\mu'_k)\}_{k=0}^{m+1}$ becomes into $\xi'_k=\xi_k$ and $\mu'_k=\mu_k$, for $k=0,\ldots,l$; $\xi'_{l+1}=\xi'$ and $\mu'_{l+1}=\mu$; and $\xi'_k=\xi_{k-1}$ and $\mu'_k=\mu_{k-1}$ for $k=l+2,\ldots,m+1$.

C.3.3 Deleting a change point.

Here we propose $\{(\xi'_k,\mu'_k)\}_{k=0}^{m-1}$, given m, which is done by drawing an index value k' uniformly from the set $\{1,\ldots,m\}$ and defining $\xi'_k=\xi_k$, for $k=0,\ldots,k'-1$, and $\xi'_k=\xi_{k+1}$, for $k=k',\ldots$. Updating the states of the random levels requires the following steps: i) For $k=0,\ldots,k'-1$, we define $\mu'_k=\mu_k$, and ii) for $k=k',\ldots$, $\mu'_k=\mu_{k+1}$ is defined. Recall that λ^* will vary at each iteration of the Gibbs sampler, so we shall update μ only in the interval $(0,\lambda^*]$, keeping the remaining components of μ beyond λ^* unaltered until the next iteration.

If in the sampler the set $\{t:d_t=d_{t-1}\}$ is empty (which is not a common case), then we will not have information concerning μ at that iteration. Hence, in the absence of statistical evidence about μ the sampler will keep the same states as in the previous iteration of the chain. However, we do not expect to observe many scenarios like this, as the likelihood shall be driven by the data driven the dynamics of the process, so in fairly few iterations or none the set $\{t:d_t=d_{t-1}\}$ will be empty.

C.4 Full conditional distribution for Q

After conditionning on $\boldsymbol{u}$, the function Q is characterized by the finite collection of latent variables $\{W_j\}_{j=1}^{u^*}$ and $\{\theta_j\}_{j=1}^{u^*}$, with $u^*=\max\{u'_t:t=1,\ldots,T\}$, for whom the full con-

ditional distributions are obtained due to the stick-breaking representation of Q. Analytic expressions are obtained though the blocked Gibbs sampler (Ishwaran and James, 2001).

C.4.1 Full conditional distribution for W_j

The W_j's that are updated depend on the set of current latent indicator variables $\{d_t\}$ and, implicitly, on the set $\{d_1^*, \ldots, d_m^*\}$ of different values of $\{d_t\}$. Thus, following the multinomial updating scheme, the full conditional distribution for the $\{W_j\}_{j=1}^{u^*}$ will be given in terms of the associated full conditional distribution of stick-breaking variables $\{V_j\}_{j=1}^{u^*}$, with the following updating scheme

$$W_1' = V_1' \quad \text{and} \quad W_j' = V_j' \prod_{l<j} (1 - V_l'),$$

where

$$V_l' \overset{\text{ind}}{\sim} \text{Be}(1 - \alpha_{\text{PDP}} + M_l, l\alpha_{\text{PDP}} + \sum_{i=l+1}^{u^*} M_i),$$

for $l = 1, \ldots, u^*$, and $M_l = \#\{d_t : d_t = l\}$.

C.4.2 Full conditional distribution for θ_j

Similarly, when conditioning on $\{u_t\}$, only a finite set of parameters $\{\theta_j\}_{j=1}^{u^*}$ need to be updated. The full conditional distribution for each latent variable is given by

$$\begin{aligned} p(\theta_j | \cdots) \quad \propto \quad & \text{Ga}(\theta_j | a_{Q_0}, b_{Q_0}) \cdot \text{We}(\lambda_1' | \theta_j) \cdot \left\{ \prod_{t: d_t' \neq d_{t-1}', d_t' = j} \text{We}(\lambda_t' | \theta_j) \right\} \\ & \cdot \left\{ \prod_{t: d_t' = d_{t-1}', d_t' = j} g_1(\lambda_t' | \lambda_{t-1}'; \mu', \theta_j) \right\} \end{aligned} \tag{C.13}$$

for $j = 1, \ldots, u^*$, with (d_t', λ_t', μ') being the current states of the Markov chain. Let us notice that the first components of the above expression coincide with the posterior kernel of the parametric Weibull-gamma model for the shape parameter θ_t. Such a distribution does not have a closed analytic form, so for that part a further sampling method will be required, e.g. the slice sampler. Regarding the θ_j's for which there is no d_t''s coinciding in value with the index j, the updated θ_j' will have to be drawn from its marginal distribution Q_0, i.e. from a $\text{Ga}(a_{Q_0}, b_{Q_0})$ in our case.

C.5 Full conditional distribution for $\boldsymbol{z}$

At each iteration of the posterior sampler the current states for $\{d_t\}$ will define a latent vector of variable dimension, $\boldsymbol{z} = (z_t : d_t = d_{t-1}, t = 2, \ldots, T)$, according to the cases where $d_t = d_{t-1}$. For all those latent variables, the full conditional distribution would be

$$\pi(z_t|\mu, \boldsymbol{z}_{-t}, \boldsymbol{\lambda}, \ldots) \propto g_{1,z_t}(\lambda_t|\lambda_{t-1}; \theta_{d_t}, \mu) p(z_t|\lambda_{t-1}; \theta_{d_t}, \mu), \tag{C.14}$$

where $g_{1,z_t}(\lambda_t|\ldots)$ and $p(z_t|\ldots)$ are given.

Hence, updating z_t at each time $t \in \{t : d_t = d_{t-1}\}$ is done according to two cases induced by the triplet (d_{t-1}, d_t, d_{t+1}): 1) $d_{t-1} = d_t$ and $d_t = d_{t+1}$, and 2) $d_{t-1} = d_t$ and $d_t \neq d_{t+1}$. Those cases will give rise to the following sampling scheme:

1) For the case $d_{t-1} = d_t$ and $d_t = d_{t+1}$, drawing z_t depend on the current states of θ_{d_t}. In terms of the λ's involved, one shall consider the three following cases: i) $\lambda'_{t-1} < \lambda_{t+1}$, ii) $\lambda'_{t-1} = \lambda_{t+1}$, and iii) $\lambda'_{t-1} > \lambda_{t+1}$. Such cases, combined with the current states of z_{t+1} and λ_t, give rise to the following sampling scheme:

   i) The case $\lambda'_{t-1} < \lambda_{t+1}$ combined with the current value of z_{t+1} gives rise to the following sampling scheme:

      a) If $z_{t+1} = 1$, then we randomly sample z'_t from the set $\{-1, 0, 1\}$ according to the distribution

      $$\pi(z_t|\mu, \boldsymbol{z}_{-t}, \boldsymbol{\lambda}, \ldots),$$

      which is described below.

      b) If $z_{t+1} = 0$, then the only admissible value for z'_t is to be equal to 1.

      c) Similarly, if $z_{t+1} = -1$, then the only admissible option for z'_t is to be equal to 1.

   ii) The case $\lambda'_{t-1} < \lambda_{t+1}$ combined with the current value of z_{t+1} gives rise to the following sampling scheme:

      a) If $z_{t+1} = 1$, then $z'_t = -1$.

      b) If $z_{t+1} = 0$, then $z'_t = 0$.

c) If $z_{t+1} = -1$, then $z'_t = 1$.

iii) Lastly, the case $\lambda'_{t-1} > \lambda_{t+1}$ combined with the current value of z_{t+1} gives rise to the following sampling scheme:

a) If $z_{t+1} = 1$, then $z'_t = -1$.

b) If $z_{t+1} = 0$, then $z'_t = -1$.

c) Similarly, if $z_{t+1} = -1$, then it is required to sample z'_t in the set $\{-1, 0, 1\}$ according to the distribution

$$\pi(z_t|\mu, z, \boldsymbol{\lambda}),$$

which is given below.

With regards to the cases where it is required to drawing a sample of z'_t from the set $\{-1, 0, 1\}$, the full conditional distribution $\pi(z_t|\mu, z, \boldsymbol{\lambda})$ will typically be defined by

$$\pi(z_t|\mu, z, \boldsymbol{\lambda}) \propto g_{1,z_t}(\lambda_t|\lambda_{t-1}; \theta, \mu)p(z_t|\lambda_{t-1}; \theta, \mu). \tag{C.15}$$

Once more, the functional form of that full conditional will be case sensitive. The cases we shall now explore are: i) $\lambda_t < \lambda_{t-1}$, ii) $\lambda_t = \lambda_{t-1}$, and iii) $\lambda_t > \lambda_{t-1}$.

i) The case $\lambda_t < \lambda_{t-1}$ gives rise to the following full conditional distribution for z'_t,

$$z'_t|\lambda_t, \lambda_{t-1}, \theta, \mu = \begin{cases} -1 & \text{with prob.} \propto g_{1,-1}(\lambda_t|\lambda_{t-1}; \theta, \mu)p(z'_t = -1|\lambda_{t-1}, \theta, \mu) \\ 0 & \text{with prob.} \propto p(z'_t = 0|\lambda_{t-1}, \theta, \mu) \\ 1 & \text{with prob.} \propto p(z'_t = 1|\lambda_{t-1}, \theta, \mu). \end{cases} \tag{C.16}$$

ii) The case $\lambda_t = \lambda_{t-1}$ gives rise to the full conditional distribution given by,

$$z'_t|\lambda_t, \lambda_{t-1}, \theta, \mu, y_t = \begin{cases} -1 & \text{with prob.} \propto p(z'_t = -1|\lambda_{t-1}, \theta, \mu) \\ 0 & \text{with prob.} \propto \mathrm{N}(y_t|0, \lambda_{t-1})p(z'_t = 0|\lambda_{t-1}, \theta, \mu) \\ 1 & \text{with prob.} \propto p(z'_t = 1|\lambda_{t-1}, \theta, \mu). \end{cases} \tag{C.17}$$

iii) The full conditional distribution corresponding to the third case, i.e. $\lambda_t > \lambda_{t-1}$, derives into,

$$z_t'|\lambda_t, \lambda_{t-1}, \theta, \mu = \begin{cases} -1 & \text{with prob.} \propto p(z_t' = -1|\lambda_{t-1}, \theta, \mu) \\ 0 & \text{with prob.} \propto p(z_t' = 0|\lambda_{t-1}, \theta, \mu) \\ 1 & \text{with prob.} \propto g_{1,1}(\lambda_t|\lambda_{t-1}; \theta, \mu)p(z_t' = 1|\lambda_{t-1}, \theta, \mu). \end{cases} \tag{C.18}$$

And such expressions are valid for $t = 2, \ldots, T$. Notice that in all the above expressions, the functions $p(z|\lambda, \theta, \mu)$ and $g_{1,z}(\lambda_t|\lambda_{t-1}; \mu, \theta)$ are defined accordingly to the definitions we have used so far.

2) Updating z_t in the case of having $d'_{t-1} = d'_t$ and $d'_t \neq d'_{t+1}$ can be done by calculating the probabilities given above, which depend on the current parameter θ_{d_t}.

C.6 Full conditional distribution for $\boldsymbol{\lambda}$

The full conditional distribution for $\boldsymbol{\lambda}$ is characterised by the individual full conditionals for λ_t, whose specification and updating scheme will now be case sensitive according to different combinations induced by current latent index variables (d_{t-1}, d_t, d_{t+1}): I) $d_{t-1} \neq d_t$ and II) $d_{t-1} = d'_t$, for $t = 2, \ldots, T$. The full conditionals and updating sampling scheme induced by those cases can be spread into:

I) For the case $d_{t-1} \neq d_t$, full conditional distribution for λ_t will be given by

$$p(\lambda_t|y_t; \theta_{d_t}) \propto \mathrm{N}(y_t|0, \lambda_t)\,\mathrm{We}(\lambda_t|\theta_{d_t}).$$

Updating λ_t can be done using the slice sampler.

II) The case case $d_{t-1} = d_t$ will be case sensitive to the additional relationship between d_t and d_{t+1}. Hence, updating λ_t in this case can be done according to:

IIa) The case of having $d_t \neq d_{t+1}$, for which the updated λ_t can be obtained from the conditional distribution

$$p(\lambda_t|y_t, z_t, \lambda_{t-1}; \theta_{d_t}) \propto \mathrm{N}(y_t|0, \lambda_t)g_{1,z_t}(\lambda_t|\lambda_{t-1}; \mu, \theta_{d_t}),$$

where $g_{1,z_t}(\lambda_t|\lambda_{t-1}; \mu, \theta_{d_t})$ is already given.

IIb) The case of having $d_t = d_{t+1}$ will induce the following full conditional distribution for λ_t,

$$p(\lambda_t|y_t, z_t, z_{t+1}, \lambda_{t-1}, \lambda_{t+1}; \theta_{d_t}) \propto \mathrm{N}(y_t|0, \lambda_t) g_{1,z_t}(\lambda_t|\lambda_{t-1}; \mu, \theta_{d_t}) \cdot g_{1,z_{t+1}}(\lambda_{t+1}|\lambda_t; \mu, \theta_{d_t}) p(z_{t+1}|\lambda_t; \mu, \theta_{d_t}),$$

with $g_{1,z_t}(\lambda_t|\lambda_{t-1}; \mu, \theta_{d_t})$ and $p(z_{t+1}|\lambda_t; \mu, \theta_{d_t})$ also given.

In the above cases, updating λ_t will require slice sampler steps.

C.6.1 Further details on the full conditional distributions for λ_t

Due to the way each z_t is defined in the model, and that at each t the previous latent variable λ_{t-1} splits the support of λ_t, and similarly λ_t splits the support of λ_{t+1}, the full conditional distributions derived above are actually case sensitive. The cases we shall explore depend on $(\lambda'_{t-1}, z'_t, z_{t+1}, \lambda_{t+1}, y_t)$, specifically on the different combination according to the relation between the pair λ'_{t-1} and λ_{t+1}, namely: i) $\lambda'_{t-1} < \lambda_{t+1}$, ii) $\lambda'_{t-1} = \lambda_{t+1}$, and iii) $\lambda'_{t-1} > \lambda_{t+1}$. Associated with each of the above cases, we shall also need to explore the derived combinations between the values that z'_t and z_{t+1} may take.

Also, the full conditionals will depend on the time period at which the sampler will be standing. Let us start considering the full conditional distribution for the periods $t = 2, \ldots, T-1$.

i) The case $\lambda'_{t-1} < \lambda_{t+1}$ will require to explore five different sub-cases in terms of z's: a) $z'_t = -1$ and $z_{t+1} = 1$, b) $z'_t = 0$ and $z_{t+1} = 1$, c) $z'_t = 1$ and $z_{t+1} = 1$, d) $z'_t = 1$ and $z_{t+1} = 0$, and e) $z'_t = 1$ and $z_{t+1} = -1$, for which it is possible to derive the full conditional distributions:

   a) The sub-case $z'_t = -1$ and $z_{t+1} = 1$ will give rise to a full conditional distribution for λ'_t given by

$$p(\lambda'_t|\ldots) \propto \mathrm{N}(y_t|0, \lambda'_t) \frac{g_\mu(\lambda'_t|\lambda'_{t-1}) \mu_{k(\lambda'_t)} S_\mu(\lambda'_t)}{S_0(\lambda'_t)} \mathbb{1}(\lambda'_t \in (0, \lambda'_{t-1})).$$

   b) The combination $z'_t = 0$ and $z_{t+1} = 1$ will give rise to a degenerated full conditional distribution, given by,

$$p(\lambda'_t|\ldots) \propto \delta_{\lambda_{t-1}}(\lambda'_t).$$

c) The combination $z'_t = 1$ and $z_{t+1} = 1$ induces the following full conditional distribution,

$$p(\lambda'_t|\ldots) \propto \mathrm{N}(y_t|0,\lambda'_t)\frac{g_0(\lambda'_t)\mu_{k(\lambda'_t)}S_\mu(\lambda'_t)}{S_0(\lambda'_t)}\,\mathbb{1}(\lambda'_t \in (\lambda'_{t-1},\lambda_{t+1})).$$

d) The combination $z'_t = 1$ and $z_{t+1} = 0$ will give rise to a degenerated full conditional distribution, given by

$$p(\lambda'_t|\ldots) \propto \delta_{\lambda_{t+1}}(\lambda'_t).$$

e) The last sub-case corresponding to the combination $z'_t = 1$ and $z_{t+1} = -1$ will induce the following conditional distribution

$$p(\lambda'_t|\ldots) \propto \mathrm{N}(y_t|0,\lambda'_t)g_0(\lambda'_t)\,\mathbb{1}(\lambda'_t \in (\lambda_{t+1},\infty)).$$

ii) The second case on the λ's, i.e. when $\lambda'_{t-1} = \lambda_{t+1}$, requires the exploration of three sub-cases on the z's, i.e.: i) $z'_t = -1$ and $z_{t+1} = 1$, b) $z'_t = 0$ and $z_{t+1} = 0$, and iii) $z'_t = 1$ and $z_{t+1} = -1$. Let us explore them in detail assuming that $\hat{\lambda} = \lambda'_{t-1} = \lambda_{t+1}$.

a) The sub-case $z'_t = -1$ and $z_{t+1} = 1$ induces the following conditional distribution for λ'_t,

$$p(\lambda'_t|\ldots) \propto \mathrm{N}(y_t|0,\lambda'_t)\frac{g_\mu(\lambda'_t|\lambda'_{t-1})\mu_{k(\lambda'_t)}S_\mu(\lambda'_t)}{S_0(\lambda'_t)}\,\mathbb{1}(\lambda'_t \in (0,\hat{\lambda})).$$

b) The combination $z'_t = 0$ and $z_{t+1} = 0$ will induce the following degenerated distribution

$$p(\lambda'_t|\ldots) \propto \delta_{\hat{\lambda}}(\lambda'_t).$$

c) The last sub-case corresponding to (ii) is associated with the combination $z'_t = 1$ and $z_{t+1} = -1$, which will induce the following conditional distribution

$$p(\lambda'_t|\ldots) \propto \mathrm{N}(y_t|0,\lambda'_t)g_0(\lambda'_t)\,\mathbb{1}(\lambda'_t \in (\hat{\lambda},\infty)).$$

iii) As in case (i), the combination $\lambda'_{t-1} > \lambda_{t+1}$ will have associated five different sub-cases in terms of z's. That is: a) $z'_t = -1$ and $z_{t+1} = 1$, b) $z'_t = -1$ and $z_{t+1} = 0$, c) $z'_t = -1$ and $z_{t+1} = -1$, d) $z'_t = 0$ and $z_{t+1} = -1$, and e) $z'_t = 1$ and $z_{t+1} = -1$, which of them inducing the following full conditional distributions:

a) The sub-case $z'_t = -1$ and $z_{t+1} = 1$ will give rise to a full conditional distribution for λ'_t given by

$$p(\lambda'_t|\ldots) \propto \mathrm{N}(y_t|0, \lambda'_t)\frac{g_\mu(\lambda'_t|\lambda'_{t-1})\mu_{k(\lambda'_t)}S_\mu(\lambda'_t)}{S_0(\lambda'_t)}\,\mathbb{1}(\lambda'_t \in (0, \lambda_{t+1})).$$

b) The combination $z'_t = -1$ and $z_{t+1} = 0$ will give rise to degenerated distribution, given by

$$p(\lambda'_t|\ldots) \propto \delta_{\lambda_{t+1}}(\lambda'_t).$$

c) The combination $z'_t = -1$ and $z_{t+1} = -1$ will induce the following distribution

$$p(\lambda'_t|\ldots) \propto \mathrm{N}(y_t|0, \lambda'_t)g_\mu(\lambda'_t)\,\mathbb{1}(\lambda'_t \in (\lambda_{t+1}, \lambda'_{t-1})).$$

d) The combination $z'_t = 0$ and $z_{t+1} = -1$ gives rise to another degenerated distribution

$$p(\lambda'_t|\ldots) \propto \delta_{\lambda'_{t-1}}(\lambda'_t).$$

e) The final sub-case to be explored in (i) corresponds to the combination $z'_t = 1$ and $z_{t+1} = -1$, which will induce the following conditional distribution

$$p(\lambda'_t|\ldots) \propto \mathrm{N}(y_t|0, \lambda'_t)g_0(\lambda'_t)\,\mathbb{1}(\lambda'_t \in (\lambda'_{t-1}, \infty)).$$

Now, for the period $t = T$ the full conditional distribution of λ'_T is defined in terms the three cases induced by z'_T. That is: a) $z'_T = -1$, b) $z'_T = 0$, and c) $z'_T = 1$. The full conditional for those cases are defined below.

a) For the case $z'_T = -1$, the full conditional distribution for λ'_T becomes into,

$$p(\lambda'_T|\ldots) \propto \mathrm{N}(y_T|0, \lambda'_T)g_\mu(\lambda'_T|\lambda'_{T-1})\,\mathbb{1}(\lambda'_T \in (\lambda'_T \in (0, \lambda'_{T-1})).$$

b) The case $z'_T = 0$ gives rise to a degenerated distribution for λ'_T,

$$p(\lambda'_T|\ldots) \propto \delta_{\lambda'_{T-1}}(\lambda'_T).$$

c) Finally, the case $z'_T = 1$ gives rise to the following full conditional distribution for λ'_T,

$$p(\lambda'_T|\ldots) \propto \mathrm{N}(y_T|0, \lambda'_T)g_0(\lambda'_T)\,\mathbb{1}(\lambda'_T \in (\lambda'_T \in (\lambda'_{T-1}, \infty)).$$

And, at period $t = 1$, the full conditional distribution for λ_1' becomes

$$p(\lambda_1'|\ldots) \propto \mathrm{N}(y_1|0, \lambda_1')g_0(\lambda_1')\,\mathbb{1}(\lambda_1' \in (0, \infty)).$$

Let us notice that none of the non-degenerated full conditional distributions above have a closed analytic form according to their support. Thus, drawing samples from them should require additional efforts. Among different alternatives, here we adopt the slice sampler method to sampling from these distributions (see, Neal, 2003, for a general introduction to the slice sampler methods).